

# A hypothesis-testing perspective on the G-normal distribution theory

Shige Peng\* and Quan Zhou†

## Abstract

The G-normal distribution was introduced by Peng [2007] as the limiting distribution in the central limit theorem for sublinear expectation spaces. Equivalently, it can be interpreted as the solution to a stochastic control problem where we have a sequence of random variables, whose variances can be chosen based on all past information. In this note we study the tail behavior of the G-normal distribution through analyzing a nonlinear heat equation. Asymptotic results are provided so that the tail “probabilities” can be easily evaluated with high accuracy. This study also has a significant impact on the hypothesis testing theory for heteroscedastic data; we show that even if the data are generated under the null hypothesis, it is possible to cheat and attain statistical significance by sequentially manipulating the error variances of the observations.

**Keywords** heteroskedasticity; nonlinear heat equation; p-hacking; sublinear expectation; tail capacity

## 1 Introduction

The primary goal of this note is to study the asymptotic tail behavior of the G-normal distribution, providing a key result to the theory of sublinear expectation spaces developed by Peng [2008]. To statisticians, our result can be interpreted from a hypothesis-testing perspective. Suppose for heteroscedastic observations  $X_1, X_2, \dots$ , one wants to conduct a statistical test regarding their common mean. Then by manipulating their variances, the

---

\*Institute of Mathematics, Shandong University, Jinan, Shandong 250100, China.

†Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.

experimenter is able to reject the null hypothesis with probability greater than the nominal significance level when the data are actually generated under the null. This can be seen as a new type of “cheating with the data”, which in spirit is similar to the well-known “p-hacking” phenomenon<sup>1</sup> [Head et al., 2015].

As suggested by its name, G-normal distribution plays a central role in the sublinear expectation theory as normal distribution does in the classical probability theory. Indeed, it is the limiting distribution in the generalized “central limit theorem” for sublinear expectation spaces. A more detailed review of the G-normal distribution (and sublinear expectation spaces) will be given in Section 2. As noted in Fang et al. [2017], to characterize the tail behavior of the G-normal distribution, equivalently we can consider the following stochastic control problem (see also Theorem 1 and Definition 1.)

**Problem 1.** Let  $\epsilon_1, \epsilon_2, \dots$  be a sequence of i.i.d. random variables such that  $\mathbf{E}(\epsilon_i) = 0$ ,  $\mathbf{E}(\epsilon_i^2) = 1$  and  $\mathbf{E}(|\epsilon_i|^3) < \infty$ , defined on some filtered probability space  $(\Omega, \mathcal{F}, \mathbf{P}, \{\mathcal{F}_i\}_{i=0}^\infty)$  where  $\{\mathcal{F}_i\}_{i=0}^\infty$  is the natural filtration generated by  $\{\epsilon_i\}_{i=1}^\infty$ , i.e.  $\mathcal{F}_i = \sigma(\epsilon_1, \dots, \epsilon_i)$ . Let  $\Sigma(\underline{\sigma}, \bar{\sigma})$  be the collection of all predictable sequences with respect to  $\{\mathcal{F}_i\}_{i=0}^\infty$  that always take value in  $[\underline{\sigma}, \bar{\sigma}]$  where  $\underline{\sigma}, \bar{\sigma}$  are given constants ( $0 \leq \underline{\sigma} \leq \bar{\sigma} < \infty$ .) For any  $\{\sigma_i\}_{i=1}^n \in \Sigma(\underline{\sigma}, \bar{\sigma})$ , define  $X_i = \sigma_i \epsilon_i$  and  $\bar{X}_n = (X_1 + \dots + X_n)/n$ . The problem is to compute the following two functions and find the sequences  $\{\sigma_i\}_{i=1}^\infty$  that attain the corresponding supremums,

$$\begin{aligned} p_1(c; \underline{\sigma}, \bar{\sigma}) &:= \lim_{n \rightarrow \infty} \sup_{\{\sigma_i\} \in \Sigma(\underline{\sigma}, \bar{\sigma})} \mathbf{E}[\mathbb{1}(\sqrt{n}\bar{X}_n > c)], \\ p_2(c; \underline{\sigma}, \bar{\sigma}) &:= \lim_{n \rightarrow \infty} \sup_{\{\sigma_i\} \in \Sigma(\underline{\sigma}, \bar{\sigma})} \mathbf{E}[\mathbb{1}(\sqrt{n}|\bar{X}_n| > c)], \end{aligned} \tag{1}$$

where  $c \in [0, \infty)$  and  $\mathbb{1}$  denotes the indicator function.

If  $\underline{\sigma} = \bar{\sigma} = \sigma$ , the observations  $X_1, X_2, \dots$  are i.i.d. and thus by the classical central limit theorem, we have  $p_2(c) = 2\Phi(-c/\sigma) = 2p_1(c)$  where  $\Phi$  denotes the distribution function of the standard normal distribution. When  $\underline{\sigma} < \bar{\sigma}$ , the functions  $p_1$  and  $p_2$  are called tail capacities of the G-normal distribution, where “capacity” can be understood as a generalization of probability. The characterization of  $p_1$  and  $p_2$  is vital to the understanding of G-normal distribution. To evaluate  $p_1$  and  $p_2$ , we need solve a nonlinear heat equation, which is studied in Section 3. It turns out that  $p_1$  admits a closed-form expression but  $p_2$

---

<sup>1</sup>The term “p-hacking” refers to the the phenomenon that researchers may try out different data analysis methods until they obtain a p-value small enough.

does not. The main technical result of this paper is an asymptotic approximation for  $p_2$ , which is highly accurate and very easy to compute.

Now we explain how Problem 1 relates to hypothesis testing. Suppose we observe  $X_1, X_2, \dots$ , which are generated by the model given in Problem 1 and consider the null hypothesis  $H_0 : E(X_i) = 0$  for every  $i$ . When  $\underline{\sigma}$  is slightly smaller than  $\bar{\sigma}$ , both the heteroscedasticity (i.e. the fact that  $\text{Var}(X_i)$  is not a constant) and the dependence structure of the observations could be very difficult to detect; if  $\{X_i\}_{i=1}^n$  is treated as an i.i.d. sample, the null hypothesis can be tested using the t-statistic,

$$T_n(X) = \frac{\sqrt{n}\bar{X}_n}{\sqrt{s_n^2}}, \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (2)$$

For sufficiently large  $n$ ,  $T_n(X)$  can be treated as a standard normal variable and the probability of  $s_n^2$  being greater than  $\bar{\sigma}^2$  quickly decreases to zero. Hence, for a one-sided test with level  $\alpha$ , the null hypothesis would be rejected if  $\sqrt{n}\bar{X}_n > \bar{\sigma}\Phi^{-1}(1 - \alpha)$ . (One can also use the  $(1 - \alpha)\%$  quantile of the t-distribution here and our theory will apply equally.) Imagine that an experimenter is able to choose any  $\{\sigma_i\}_{i=1}^n$  from the set  $\Sigma(\underline{\sigma}, \bar{\sigma})$  (defined in Problem 1) and wants to maximize the probability of the event  $\{\sqrt{n}\bar{X}_n > \bar{\sigma}\Phi^{-1}(1 - \alpha)\}$ . Then, as will be shown in Section 4, the asymptotically optimal strategy is to simply choose either  $\sigma_i = \underline{\sigma}$  or  $\sigma_i = \bar{\sigma}$  depending on whether  $X_1 + \dots + X_{i-1}$  is greater than  $\sqrt{n}\bar{\sigma}\Phi^{-1}(1 - \alpha)$ . Further,  $p_1(\bar{\sigma}\Phi^{-1}(1 - \alpha))$  is always strictly greater than  $\alpha$  given that  $\bar{\sigma} > \underline{\sigma}$ . A similar analysis can be conducted for the two-sided test as well. Simulation studies with unknown  $\underline{\sigma}, \bar{\sigma}$  will be provided in Section 4.

We point out that in many applications, it is possible for the experimenter to affect the error variances. For example, consider an economist planning to survey individuals of different ages to study whether some variable has an effect on personal income. The errors are heteroscedastic because the income of older people tends to have a larger variance. Whether the economist deliberately surveys more younger (or older) people seems unimportant since age is included in the regression model as a confounding variable. But the result of this paper implies that this is not true if the economist decides who to survey next (in terms of age) based on previous observations.

## 2 G-normal distribution and Peng's central limit theorem

The sublinear expectation theory was motivated by capturing the model uncertainty in real-world markets [Artzner et al., 1999, Chen and Epstein, 2002] and has found applications in economics, mathematical finance and statistics [Epstein and Ji, 2014, Peng et al., 2018, Lin et al., 2016]. Concepts such as “distribution” and “independence” are redefined for a sublinear expectation space. But to make the present note easier to understand, we will present all the results using the language of classical probability theory, except the use of the terms “G-normal distribution” and “tail capacity”.

The central limit theorem for sublinear expectation spaces, first developed by Peng [2008], has been formulated in various ways. In Theorem 1 we present the version given in Rokhlin [2015] [see also Fang et al., 2017], which can be seen as a generalization of the classical central limit theorem to controlled stochastic processes. It is an immediate corollary of Peng's original central limit theorem, but translated into the language of classical probability (see Appendix B).

**Theorem 1.** *Let  $\{\epsilon_i\}_{i=1}^\infty$ ,  $\{\sigma_i\}_{i=1}^n$ ,  $\{X_i\}_{i=1}^n$  and  $\Sigma(\underline{\sigma}, \bar{\sigma})$  be as given in Problem 1. Then for any Lipschitz function  $\varphi$ ,*

$$\lim_{n \rightarrow \infty} \sup_{\{\sigma_i\} \in \Sigma(\underline{\sigma}, \bar{\sigma})} \mathbf{E} \left[ \varphi(\sqrt{n} \bar{X}_n) \right] = u(1, 0; \varphi), \quad (3)$$

where  $\{u(t, x; \varphi): (t, x) \in [0, \infty) \times \mathbb{R}\}$  is the unique viscosity solution to the Cauchy problem,

$$u_t = \frac{1}{2} (\bar{\sigma}^2 (u_{xx})^+ - \underline{\sigma}^2 (u_{xx})^-), \quad u(0, x) = \varphi(x). \quad (4)$$

In the above expression,  $u_t = \partial u / \partial t$ ,  $u_{xx} = \partial^2 u / \partial x^2$ , and the superscripts  $+$  and  $-$  denote the positive and negative parts respectively.

**Remark 1.** See Peng [2008], Peng [2019], Rokhlin [2015] or Fang et al. [2017, Theorem 4.1] for the proof. The theorem can be further generalized to non-identically distributed sequence  $\{\epsilon_i\}_{i=1}^\infty$ . For the convergence rate of Theorem 1, see Fang et al. [2017], Song [2019]<sup>2</sup>, Krylov [2019] and Huang and Liang [2019].

**Remark 2.** The function  $\varphi$  can also be any Borel-measurable indicator function [Peng, 2010, Peng et al., 2018].

---

<sup>2</sup>The paper of Song [2019] was submitted in 2017 and the earliest among the four.

**Remark 3.** The classical central limit theorem corresponds to the special case  $\underline{\sigma} = \bar{\sigma} = \sigma$ , where (4) reduces to the heat equation and thus

$$u(t, x; \varphi) = \int_{-\infty}^{\infty} \phi(z) \varphi(x + \sigma\sqrt{t}z) dz$$

where  $\phi(x)$  is the density function of the standard normal distribution. We call the nonlinear partial differential equation (4) a G-heat equation.

Now we are ready to define G-normal distribution. Note that G-normal distribution is not a distribution in the traditional sense, and a “random variable”  $Z$  following G-normal distribution actually has distributional uncertainty.

**Definition 1** (G-normal distribution). *Let  $\mathcal{P}_Z$  be a set of probability measures defined on the space  $(\Omega, \mathcal{F})$ . A measurable function  $Z: \Omega \mapsto \mathbb{R}$  is said to follow a G-normal distribution with lower variance  $\underline{\sigma}^2$  and upper variance  $\bar{\sigma}^2$  ( $0 \leq \underline{\sigma} \leq \bar{\sigma}$ ), if, for every Lipschitz function  $\varphi$ ,*

$$\sup_{P \in \mathcal{P}_Z} \mathbb{E}_P[\varphi(Z)] = \sup_{P \in \mathcal{P}_Z} \int_{\Omega} \varphi(Z) dP = u(1, 0; \varphi),$$

where  $u(1, 0; \varphi)$  is as given in Theorem 1.

**Remark 4.** As expected, when  $\underline{\sigma} = \bar{\sigma} = \sigma$ , the G-normal distribution reduces to the normal distribution  $N(0, \sigma^2)$ .

**Remark 5.** In the sublinear expectation theory,  $\mathcal{P}_Z$  represents the collection of all possible probability measures underlying  $Z$ , and  $u(1, 0; \varphi)$  is called the sublinear expectation of  $\varphi(Z)$ . For the existence of G-normal distribution, see Peng [2010, Theorem 2.1 and Chapter II.2].

When  $\varphi$  in Definition 1 is an indicator function, we have

$$p(A; \underline{\sigma}, \bar{\sigma}) := \sup_{P \in \mathcal{P}_Z} \int_{\Omega} \mathbb{1}(Z \in A) dP = \sup_{P \in \mathcal{P}_Z} P\{Z \in A\}.$$

Here  $p(A) = p(A; \underline{\sigma}, \bar{\sigma})$  is a set function defined for each  $A \in \mathcal{F}$ , which clearly satisfies  $p(\emptyset) = 0$  and  $p(\Omega) = 1$ . But unlike a probability measure,  $p(\cdot)$  is not an additive function; it is called a Choquet capacity [Choquet, 1954], or capacity for short [see also Chen et al., 2005, Denis et al., 2011]. In this work, we are interested in the “tail capacities” of the G-normal distribution, i.e. the functions  $p_1$  and  $p_2$  defined in (1). They are the solutions

to the G-heat equations (4) with initial condition  $\varphi(x) = \mathbb{1}(x > c)$  and  $\varphi(x) = \mathbb{1}(|x| > c)$  respectively. Understanding the tail behavior of the G-normal distribution is crucial to the asymptotic theory of sublinear expectation spaces.

For a formal exposition of the sublinear expectation theory, readers are referred to Peng [2010]. For the purposes of this paper, we only need to use Theorem 1 to compute the tail capacities of the G-normal distribution.

### 3 Tail capacities of the G-normal distribution

By Theorem 1 and Definition 1, to compute the tail capacities of the G-normal distribution, we need to solve the corresponding Cauchy problems given in (4), which is often difficult since the G-heat equation is nonlinear. Fortunately, we have a closed-form solution for the one-sided tail capacity,  $p_1(c)$ . But for the two-sided tail capacity  $p_2(c)$ , a closed-form solution is not available and we will offer an asymptotic approximation which has remarkable accuracy for large values of  $c$ . Recall that  $\phi$  and  $\Phi$  respectively denote the probability density function and the cumulative distribution function of the standard normal distribution.

#### 3.1 One-sided tail capacity

We first derive a closed-form solution to the G-heat equation (4) with initial condition  $u(0, x) = \mathbb{1}(x > c)$ . This result will be used later for approximating the solution to the G-heat equation with initial condition  $u(0, x) = \mathbb{1}(|x| > c)$ .

**Theorem 2.** *Consider the following Cauchy problem for the G-heat equation,*

$$u_t = \frac{1}{2} (\bar{\sigma}^2 (u_{xx})^+ - \underline{\sigma}^2 (u_{xx})^-), \quad u(0, x) = \mathbb{1}(x > c),$$

where  $c \in \mathbb{R}$ . The solution is given by  $u(t, x) = f\{(x - c)/\sqrt{t}\}$  where

$$f(y) = f(y; \underline{\sigma}, \bar{\sigma}) = \frac{2}{\bar{\sigma} + \underline{\sigma}} \int_{-y}^{\infty} \{\phi(z/\bar{\sigma})\mathbb{1}(z \geq 0) + \phi(z/\underline{\sigma})\mathbb{1}(z < 0)\} dz.$$

*Proof.* This result is also given in Peng et al. [2018] but without derivation. Here we show how the solution is derived. For the uniqueness and the existence of this solution, see Peng [2010], Peng et al. [2018].

Notice that if  $u(t, x)$  is the solution to this Cauchy problem, then, for any  $a > 0$ ,  $u(at, \sqrt{a}(x - c) + c)$  is also a solution. This implies that  $u(t, x) = f\{(x - c)/\sqrt{t}\}$  for some function  $f$ . Then routine calculations show that the function  $f = f(y)$  must satisfy

$$-yf_y = \bar{\sigma}^2(f_{yy})^+ - \underline{\sigma}^2(f_{yy})^-.$$

If  $u$  is twice differentiable in  $x$ ,  $f(y)$  must be continuous at 0 and thus it can be expressed as

$$f(y) = \int_{-y}^{\infty} \left\{ \frac{A_1}{\bar{\sigma}} \phi(z/\bar{\sigma}) \mathbb{1}(z \geq 0) + \frac{A_2}{\underline{\sigma}} \phi(z/\underline{\sigma}) \mathbb{1}(z < 0) \right\} dz + B,$$

where  $A_1, A_2, B_1, B_2$  are some constants to be determined. By the initial conditions,  $f(\infty) = 1$  and  $f(-\infty) = 0$ , we obtain that  $A_1 + A_2 = 2$  and  $B = 0$ . Finally, the twice differentiability of  $f(y)$  amounts to matching the left and right first derivatives at  $y = 0$ , which yields  $A_1/A_2 = \bar{\sigma}/\underline{\sigma}$ . The theorem is then proved by checking that  $u(t, x) = f\{(x - c)/\sqrt{t}\}$  is in  $C^{1,2}$ .  $\square$

**Corollary 1.** *The one-sided tail capacity of the G-normal distribution is*

$$p_1(c; \underline{\sigma}, \bar{\sigma}) = \frac{2}{\bar{\sigma} + \underline{\sigma}} \int_c^{\infty} \{ \phi(z/\bar{\sigma}) \mathbb{1}(z \geq 0) + \phi(z/\underline{\sigma}) \mathbb{1}(z < 0) \} dz.$$

*Proof.* This is immediate from the definition of G-normal distribution.  $\square$

**Remark 6.** One can check that  $p_1(c)$  behaves just like a tail probability function in the sense that, for any  $0 < \underline{\sigma} \leq \bar{\sigma}$ ,  $p_1(c)$  is monotone decreasing with  $p_1(-\infty) = 1$  and  $p_1(\infty) = 0$ .

**Corollary 2.** *The solution to the following Cauchy problem,*

$$v_t = \frac{1}{2} (\bar{\sigma}^2(v_{xx})^+ - \underline{\sigma}^2(v_{xx})^-), \quad v(0, x) = \mathbb{1}(x < -c),$$

*is given by  $v(t, x) = f\{(-x - c)/\sqrt{t}\}$  where  $f$  is as given in Theorem 2.*

*Proof.* This follows immediately from Theorem 2 by symmetry.  $\square$

### 3.2 Two-sided tail capacity

To compute the two-sided tail capacity  $p_2(c)$  defined in (1), we need to solve the G-heat equation (4) with initial condition  $u(0, x) = \mathbb{1}(|x| > c)$ , which does not admit a closed-form solution. However, we do have an asymptotic result that turns out to be very useful.

**Theorem 3.** Let  $w(t, x)$  be the solution to the Cauchy problem,

$$w_t = \frac{1}{2} (\bar{\sigma}^2 (w_{xx})^+ - \underline{\sigma}^2 (w_{xx})^-), \quad w(0, x) = \mathbb{1}(|x| > c).$$

Let  $u(t, x) = f\{(x - c)/\sqrt{t}\}$  be as given in Theorem 2 and  $v(t, x) = f\{(-x - c)/\sqrt{t}\}$  be as given in Corollary 2. For any  $x \in \mathbb{R}$  and  $t \geq 0$ , if  $c > \bar{\sigma}\sqrt{t}/2$ ,

$$0 \leq u(t, x) + v(t, x) - w(t, x) \leq \frac{2(\bar{\sigma} - \underline{\sigma})}{\bar{\sigma}} \Phi\left(-\frac{2c}{\bar{\sigma}\sqrt{t}}\right).$$

*Proof.* The goal is to show that for any fixed  $(t, x) \in [0, 1] \times \mathbb{R}$  and sufficiently large  $c$ , we have  $w(t, x) \approx u(t, x) + v(t, x)$  (it may be helpful to think of the heat transfer model to gain some intuition.) To this end, define

$$G(x) = \frac{1}{2} (\bar{\sigma}^2 x^+ - \underline{\sigma}^2 x^-), \quad \delta(t, x) = G(u_{xx}) + G(v_{xx}) - G(u_{xx} + v_{xx}).$$

Observe that  $G$  is a sublinear function and thus  $\delta(t, x) \geq 0$ . Further,  $u + v$  satisfies the initial condition that  $u(0, x) + v(0, x) = \mathbb{1}(|x| > c)$ . Now if we can find a function  $\bar{\delta}(t)$  such that  $\delta(t, x) \leq \bar{\delta}(t)$  for every  $x$ , we can apply the comparison theorem given in Peng [2010, Appendix C] to get

$$0 \leq u(t, x) + v(t, x) - w(t, x) \leq \int_0^t \bar{\delta}(s) ds, \quad (5)$$

for all  $t \geq 0, x \in \mathbb{R}$ . It is straightforward to check that

$$\delta(t, x) = \begin{cases} \frac{\bar{\sigma}^2 - \underline{\sigma}^2}{2} \min\{|u_{xx}|, |v_{xx}|\}, & \text{if } u_{xx}v_{xx} < 0, \\ 0, & \text{if } u_{xx}v_{xx} \geq 0. \end{cases} \quad (6)$$

Hence we only need to bound  $\min\{|u_{xx}|, |v_{xx}|\}$ . Direct calculations give that

$$\begin{aligned} u_{xx}(t, x) &= \frac{1}{t} f_{yy} \left( \frac{x - c}{\sqrt{t}} \right), & v_{xx}(t, x) &= \frac{1}{t} f_{yy} \left( \frac{-x - c}{\sqrt{t}} \right), \\ f_{yy}(y) &= \frac{-2y}{\bar{\sigma} + \underline{\sigma}} \left\{ \frac{1}{\bar{\sigma}^2} \phi(y/\bar{\sigma}) \mathbb{1}(y \leq 0) + \frac{1}{\underline{\sigma}^2} \phi(y/\underline{\sigma}) \mathbb{1}(y > 0) \right\}. \end{aligned}$$

Notice that we only need to bound  $\min\{|u_{xx}|, |v_{xx}|\}$  on the region where  $u_{xx}v_{xx} < 0$ , which is  $(c, \infty) \cup (-\infty, -c)$ . For  $x > c$ , we have

$$\min\{|u_{xx}|, |v_{xx}|\} \leq |v_{xx}| \leq \sup_{x > c} v_{xx}(t, x) = \frac{1}{t} \sup_{x > c} f_{yy} \left( \frac{-x - c}{\sqrt{t}} \right). \quad (7)$$

Assume  $c > \bar{\sigma}\sqrt{t}/2$ . Then for any  $x > c$ ,  $(-x-c)/\sqrt{t} < -\bar{\sigma}$ . Observing that  $\sup_{y<0} |f_{yy}(y)|$  is attained at  $y = -\bar{\sigma}$  and  $|f_{yy}(y)|$  is monotone increasing on  $(-\infty, -\bar{\sigma})$ , we can bound the right-hand side of (7) by

$$\min\{|u_{xx}|, |v_{xx}|\} \leq \frac{1}{t} f_{yy} \left( \frac{-2c}{\sqrt{t}} \right). \quad (8)$$

The same argument shows that this bound also holds for  $x < -c$  and thus holds for  $x \in \mathbb{R}$ . Using (6) and (8) and the assumption  $c > \bar{\sigma}\sqrt{t}/2$ , we obtain the bound

$$\bar{\delta}(s) = \frac{\bar{\sigma}^2 - \underline{\sigma}^2}{2s} f_{yy} \left( \frac{-2c}{\sqrt{s}} \right) = \frac{2c(\bar{\sigma} - \underline{\sigma})}{\bar{\sigma}^2 \sqrt{s^3}} \phi \left( \frac{2c}{\bar{\sigma}\sqrt{s}} \right)$$

for any  $s \in [0, t]$ . Integrating  $\bar{\delta}(s)$  with respect to  $s$  gives

$$\int_0^t \bar{\delta}(s) ds = \frac{2(\bar{\sigma} - \underline{\sigma})}{\bar{\sigma}} \Phi \left( -\frac{2c}{\bar{\sigma}\sqrt{t}} \right). \quad (9)$$

The theorem is then proved by recalling (5).  $\square$

Theorem 3 suggests that we may approximate the two-sided tail capacity using the sum of two one-sided tail capacities. For any fixed  $t$ , as  $c \rightarrow \infty$ , the error term (9) quickly goes to zero due to the fast decay of the function  $\Phi$ . However, for any  $(t, x)$ ,  $u(t, x) + v(t, x)$  also goes to zero as  $c \rightarrow \infty$ . The next corollary confirms that the relative error of this approximation is negligible.

**Corollary 3.** *Under the setup of Theorem 3 and assuming  $c > \bar{\sigma}/2$ , we have*

$$0 \leq u(t, x) + v(t, x) - w(t, x) \leq \frac{(\bar{\sigma} - \underline{\sigma})\sqrt{t}}{c\sqrt{2\pi}} \exp \left( -\frac{2c^2}{\bar{\sigma}^2 t} \right).$$

Further, the relative error is given by

$$0 \leq \frac{u(t, x) + v(t, x) - w(t, x)}{u(t, x) + v(t, x)} \leq \frac{(\bar{\sigma}^2 - \underline{\sigma}^2)(c^2/\bar{\sigma}^2 + t)}{4c^2} \exp \left( -\frac{3c^2}{2\bar{\sigma}^2 t} \right).$$

*Proof.* It is well known that the tail probabilities of the standard normal distribution can be bounded by

$$\frac{x}{x^2 + 1} \leq \frac{\Phi(-x)}{\phi(x)} \leq \frac{1}{x}.$$

Hence the error term (9) can be bounded by

$$\int_0^t \bar{\delta}(s) ds \leq \frac{(\bar{\sigma} - \underline{\sigma})\sqrt{t}}{c\sqrt{2\pi}} \exp \left( -\frac{2c^2}{\bar{\sigma}^2 t} \right).$$

By straightforward calculations, one can show that  $u(t, x) + v(t, x)$  is minimized at  $x = 0$ . By Theorem 2, for  $c \geq 0$ ,

$$u(t, 0) = v(t, 0) = \frac{2\bar{\sigma}}{\bar{\sigma} + \underline{\sigma}} \Phi\left(-\frac{c}{\bar{\sigma}\sqrt{t}}\right) \geq \frac{\sqrt{2}}{\sqrt{\pi}(\bar{\sigma} + \underline{\sigma})} \frac{c\bar{\sigma}^2\sqrt{t}}{c^2 + \bar{\sigma}^2 t} \exp\left(-\frac{c^2}{2\bar{\sigma}^2 t}\right),$$

which then yields the relative error bound stated in the theorem.  $\square$

**Corollary 4.** *Let  $p_1, p_2$  be the tail capacities of the G-normal distribution as given in (1). As  $c \rightarrow \infty$ , we have  $p_2(c) \sim 2p_1(c)$  with the relative error given by*

$$\frac{|p_2(c) - 2p_1(c)|}{2p_1(c)} \lesssim \frac{1 - \underline{\sigma}^2/\bar{\sigma}^2}{4} \exp\left(-\frac{3c^2}{2\bar{\sigma}^2}\right),$$

where  $\sim$  denotes asymptotic equivalence and  $\lesssim$  means “asymptotically less than”.

*Proof.* Notice that, by Definition 1,  $2p_1(c) = u(1, 0) + v(1, 0)$  and  $p_2(c) = w(1, 0)$ . Hence we only need to apply Corollary 3 with  $t = 1$ .  $\square$

**Remark 7.** This approximation is accurate enough for usual purposes. For example, using the bound given in Corollary 3, for  $\bar{\sigma} = 1, \underline{\sigma} = 0.8$ , we have  $p_2(\Phi^{-1}(0.95)) = 0.11$  with RE (relative error)  $< 0.002$ ,  $p_2(\Phi^{-1}(0.975)) = 0.056$  with RE  $< 4 \times 10^{-4}$ , and  $p_2(\Phi^{-1}(0.995)) = 0.011$  with RE  $< 5 \times 10^{-6}$ .

## 4 Hypothesis testing with heteroscedastic data

To completely solve Problem 1, we need to find the asymptotically optimal policies  $\{\sigma_i\}_{i=1}^{\infty}$  that attain the supremums in (1). This can be most conveniently computed using the stochastic control theory and Hamilton-Jacobi-Bellman equation, which we briefly explain in Appendix A.

Now we come back to the hypothesis testing problem described in Section 1. First, consider the one-sided test. Since all the observations  $X_1, X_2, \dots$  have mean zero and variance less than or equal to  $\bar{\sigma}^2$ , as  $n \rightarrow \infty$ ,  $s_n^2$  is less than or equal to  $\bar{\sigma}^2$  with probability one. Hence, if we conduct a test with rejection region  $\{T_n(X) > \Phi^{-1}(1 - \alpha)\}$ , the probability of rejecting the null hypothesis,  $E(X_i) = 0$ , can be bounded from below by  $P(\sqrt{n}\bar{X}_n > \bar{\sigma}\Phi^{-1}(1 - \alpha))$ . However, if the experimenter is able to choose any  $\{\sigma_i\} \in \Sigma(\underline{\sigma}, \bar{\sigma})$ , then

$$\lim_{n \rightarrow \infty} \sup_{\{\sigma_i\} \in \Sigma(\underline{\sigma}, \bar{\sigma})} P(\sqrt{n}\bar{X}_n > \bar{\sigma}\Phi^{-1}(1 - \alpha)) = p_1(\bar{\sigma}\Phi^{-1}(1 - \alpha); \underline{\sigma}, \bar{\sigma}), \quad (10)$$

and by Theorem 2, for any  $\alpha \leq 0.5$  and  $\underline{\sigma} < \bar{\sigma}$ ,

$$p_1(\bar{\sigma}\Phi^{-1}(1 - \alpha); \underline{\sigma}, \bar{\sigma}) = \frac{2\alpha}{1 + \underline{\sigma}/\bar{\sigma}} > \alpha. \quad (11)$$

Note that, if  $X_1, X_2, \dots$  are independent, the test is still asymptotically valid in the sense that the type I error rate is  $\alpha$  as  $n \rightarrow \infty$ . By violating the independence assumption and carefully manipulating the sequence  $\{\sigma_i\}_{i=1}^\infty$ , the experimenter is able to increase the probability of falsely rejecting the null hypothesis by at least  $2\bar{\sigma}/(\underline{\sigma} + \bar{\sigma})$ . By (A3) and Theorem 2, one can show that the asymptotically optimal strategy to attain the supremum in (10) is

$$\begin{aligned} \sigma_i &= \bar{\sigma}, & \text{if } S_{i-1}/\sqrt{n} \leq \bar{\sigma}\Phi^{-1}(1 - \alpha), \\ \sigma_i &= \underline{\sigma}, & \text{if } S_{i-1}/\sqrt{n} > \bar{\sigma}\Phi^{-1}(1 - \alpha), \end{aligned} \quad (12)$$

where  $S_i = X_1 + \dots + X_i$ .

The two-sided test can be analyzed similarly. The probability of rejecting the null hypothesis can be bounded from below by  $P(\sqrt{n}|\bar{X}_n| > \bar{\sigma}\Phi^{-1}(1 - \alpha/2))$ . By Corollary 4, if  $\underline{\sigma} < \bar{\sigma}$  and  $\alpha$  is small,

$$p_2(\bar{\sigma}\Phi^{-1}(1 - \alpha/2); \underline{\sigma}, \bar{\sigma}) \approx 2p_1(\bar{\sigma}\Phi^{-1}(1 - \alpha/2); \underline{\sigma}, \bar{\sigma}) = \frac{2\alpha}{1 + \underline{\sigma}/\bar{\sigma}} > \alpha.$$

The control policy given in (A3) can still be expressed in the form like (12); however,  $S_i$  is replaced by  $|S_i|$  and the threshold changes with  $i$ . Numerically we can compute the threshold for each  $i$  by solving  $w_{xx} = 0$  where  $w$  is as in Theorem 3. We observe that, for small  $\alpha$ , the threshold goes to  $\Phi^{-1}(1 - \alpha/2)$  very quickly as  $i$  increases and thus can be treated just as a constant.

Nevertheless, for a finite sample size, we need to take into account the distribution of the sample variance  $s_n^2$ , and the Z-test should be replaced by the t-test. More importantly, in practice the parameters  $\underline{\sigma}$  and  $\bar{\sigma}$  are usually unknown. Inspired by our theoretical results, we propose the following heuristic control policy for a two-sided test,

$$\begin{aligned} \sigma_i &= \bar{\sigma}, & \text{if } |S_{i-1}|/\sqrt{ns_{i-1}^2} \leq c_\alpha, \\ \sigma_i &= \underline{\sigma}, & \text{if } |S_{i-1}|/\sqrt{ns_{i-1}^2} > c_\alpha. \end{aligned} \quad (13)$$

where  $c_\alpha$  is the critical value and  $s_i^2$  is the sample variance of  $X_1, \dots, X_i$ .

We end our paper with a realistic simulation study. Consider a two-sided t-test with  $\alpha = 0.05$ . We generate normal samples  $X_1, \dots, X_n$  (with mean zero and variances  $\sigma_1^2, \dots, \sigma_n^2$ ) using (13) with  $\bar{\sigma} = 1$  and  $\underline{\sigma} = 0.8$ . Then we compute the t-statistic by (2)

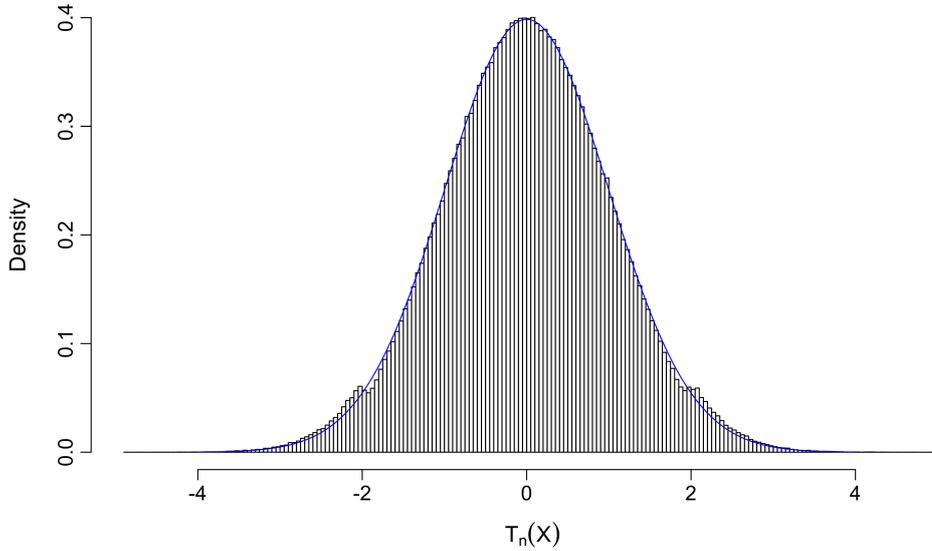


Figure 1: Histogram of the empirical distribution of the t-statistic  $T_n(X)$  for  $n = 200$ . The normal observations  $X_1, \dots, X_n$  are generated sequentially using (13). The blue curve represents the density function of the t-distribution with  $n - 1$  degrees of freedom. For  $n = 200$  and  $\alpha = 0.05$ , the critical values are  $\pm 1.97$ .

(assuming  $\underline{\sigma}, \bar{\sigma}$  are unknown) and perform a two-sided t-test with degree of freedom equal to  $n - 1$ . We repeat this experiment for 1 million times. For  $n = 20$ , the null hypothesis  $E(X_i) = 0$  is rejected in 5.65 percent of all the experiments; for  $n = 200$ , the null is rejected in 5.89 percent of all the experiments. To further illustrate how the manipulation of the variances of  $X_1, \dots, X_n$  affects the sampling distribution of the t-statistic, we compare the empirical distribution of  $T_n(X)$  with t-distribution in Figure 1 for  $n = 200$ . It can be seen that the empirical distribution of  $T_n(X)$  almost coincides with the theoretical t-distribution, except near the critical values (approximately  $\pm 2$ ).

In the simulation we use  $\bar{\sigma} = 1$  and  $\underline{\sigma} = 0.8$  to reflect that in reality the possible influence from the experimenter is limited. Because  $\underline{\sigma}, \bar{\sigma}$  are close, every simulated set of observations  $\{X_i\}_{i=1}^n$  looks just like a homoscedastic normal sample. Further, without prior knowledge, the dependence structure is almost impossible to detect. But compared to the nominal significance level  $\alpha = 0.05$ , the type I error rate is inflated by 13% for  $n = 20$  and 18% for  $n = 200$ .

## Appendix A

The asymptotically optimal policy  $\{\sigma_i\}_{i=1}^{\infty}$  that attains the supremums in (1) can be derived using the diffusion limit of Problem 1. Let  $W_t$  be a Wiener process and  $Y_t$  be a controlled process evolving by  $dY_t = \sigma_t dW_t$ , where  $\sigma_t$  is a process progressively measurable with respect to the natural filtration generated by  $W_t$ . Let  $S_i = X_1 + \dots + X_i$  where  $\{X_i\}_{i=1}^{\infty}$  are as defined in Problem 1. Intuitively speaking, as  $n \rightarrow \infty$ ,  $S_i/\sqrt{n}$  has the same distribution as  $Y_{i/n}$ , and in particular,  $\sqrt{n}\bar{X}_n$  has the same distribution as  $Y_1$  (recall Donsker's Theorem.)

Consider the value function,

$$V(t, y; \varphi) = \sup_{\sigma_t \in [\underline{\sigma}, \bar{\sigma}]} \mathbf{E}[\varphi(Y_1) \mid Y_t = y], \quad (t, y) \in [0, 1] \times \mathbb{R},$$

where the supremum is taken over all the progressively measurable processes that take value in  $[\underline{\sigma}, \bar{\sigma}]$  on the time interval  $[0, 1]$ . In our case,  $\varphi$  is given by  $\varphi(y) = \mathbb{1}(y > c)$  or  $\varphi(y) = \mathbb{1}(|y| > c)$ . It is not difficult to see that the optimal control  $\sigma_t$  must be a measurable function of  $(t, Y_t)$  by a Markovian argument. Then we may guess the solution by solving the Hamilton-Jacobi-Bellman (HJB) equation,

$$\sup_{\sigma_t \in [\underline{\sigma}, \bar{\sigma}]} \left\{ \frac{\partial V}{\partial t} + \frac{\sigma_t^2}{2} \frac{\partial^2 V}{\partial y^2} \right\} = 0, \quad V(0, y) = \varphi(y), \quad (\text{A1})$$

and then prove it using the so-called verification techniques [for more details see, for example, Yong and Zhou, 1999]. As expected, the HJB approach yields the same result as Theorem 1, and indeed, the solution to (A1) is given by

$$V(t, y; \varphi) = u(1 - t, y; \varphi), \quad (\text{A2})$$

where  $u$  is the solution to the G-heat equation (4).

The optimal control process  $\sigma_t$  is the one which attains the supremum in (A1); that is,

$$\begin{aligned} \sigma_t &= \bar{\sigma}, & \text{if } V_{yy}(t, Y_t) \geq 0, \\ \sigma_t &= \underline{\sigma}, & \text{if } V_{yy}(t, Y_t) < 0. \end{aligned}$$

Using (A2), the corresponding discrete-time approximation is given by

$$\begin{aligned} \sigma_i &= \bar{\sigma}, & \text{if } u_{xx}(1 - (i - 1)/n, S_{i-1}/\sqrt{n}) \geq 0, \\ \sigma_i &= \underline{\sigma}, & \text{if } u_{xx}(1 - (i - 1)/n, S_{i-1}/\sqrt{n}) < 0, \end{aligned} \quad (\text{A3})$$

for  $i = 1, \dots, n$ .

## Appendix B

Here we provide a brief presentation of Peng's central limit theorem [Peng, 2008, 2019] and show that it immediately implies Theorem 1. We refer the readers to Peng [2010] for further details. Let  $\mathcal{H}$  be a linear space of real-valued functions defined on a set  $\Omega$  such that if  $X_1, \dots, X_n \in \mathcal{H}$ , we have  $\varphi(X_1, \dots, X_n) \in \mathcal{H}$  for any Lipschitz function  $\varphi$ . Given a collection of probability measures  $\mathcal{P}$ , we can define a sublinear expectation (of some random variable  $X$ ), denoted by  $\mathbb{E}$ , by

$$\mathbb{E}[X] := \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[X] = \sup_{\mathbb{P} \in \mathcal{P}} \int_{\Omega} X d\mathbb{P}.$$

One can check that, for  $X, Y \in \mathcal{H}$  with  $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$ , we have (i)  $\mathbb{E}[X] \geq \mathbb{E}[Y]$  if  $X \geq Y$ ; (ii)  $\mathbb{E}[X] - \mathbb{E}[-Y] \leq \mathbb{E}[X + Y] \leq \mathbb{E}[X] + \mathbb{E}[Y]$ ; (iii)  $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$  for  $\lambda \geq 0$ ; (iv)  $\mathbb{E}[X + c] = \mathbb{E}[X] + c$  for  $c \in \mathbb{R}$ .

In the sublinear expectation theory, two random variables,  $X, Y \in \mathcal{H}$ , are called identically distributed (under  $\mathbb{E}$ ) iff  $\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)]$  for every Lipschitz function  $\varphi$ . For  $X_1, \dots, X_n \in \mathcal{H}$ ,  $X_n$  is said to be independent from  $Y = (X_1, \dots, X_{n-1})$  (under  $\mathbb{E}$ ) iff  $\mathbb{E}[\varphi(X_n, Y)] = \mathbb{E}[\mathbb{E}[\varphi(X_n, y)]_{y=Y}]$  for every Lipschitz  $\varphi$ . We say  $X_1, X_2, \dots$ , are i.i.d. if they are identically distributed and  $X_n$  is independent from  $X_1, \dots, X_{n-1}$  for each  $n$ .

**Theorem 4** (Peng's central limit theorem). *Let  $X_1, X_2, \dots$  be an i.i.d. sequence of random variables under sublinear expectation  $\mathbb{E}$ . If  $\mathbb{E}|X_1|^3 < \infty$ ,  $\mathbb{E}[X_1] = \mathbb{E}[-X_1] = 0$ ,  $\mathbb{E}[X_1^2] = \bar{\sigma}^2$  and  $-\mathbb{E}[-X_1^2] = \underline{\sigma}^2$  for some  $0 < \underline{\sigma} \leq \bar{\sigma} < \infty$ , then  $\{X_i\}_{i=1}^{\infty}$  converges "in distribution" to the G-normal distribution; that is, for any Lipschitz function  $\varphi$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{E} [\varphi(\sqrt{n} \bar{X}_n)] = \mathbb{E}[\varphi(Z)],$$

where  $Z$  is G-normally distributed with lower variance  $\underline{\sigma}^2$  and upper variance  $\bar{\sigma}^2$ .

As mentioned in Remark 5,  $\mathbb{E}[\varphi(Z)]$  can be computed by solving the corresponding G-heat equation. To see that the above result immediately implies Theorem 1, we only need to find an appropriate sublinear expectation space and check the conditions.

To this end, let  $\Omega := \{\omega = (\omega_i)_{i=1}^{\infty}, \omega_i \in \mathbb{R}\}$ , and

$$\mathcal{H} := \{\xi(\omega) = \varphi(\omega_1, \dots, \omega_n), \omega \in \Omega, n \in \mathbb{N}, \varphi \in C_{Lip}(\mathbb{R}^n)\}.$$

Define  $X_i(\omega) := \omega_i$  for  $i = 1, 2, \dots$ , which we call the canonical processes of  $(\Omega, \mathcal{H})$ . Clearly  $X_i(\omega) \in \mathcal{H}$ . For any random variable  $\xi(\omega) = \varphi(X_1, \dots, X_n)$  where  $\varphi \in C_{Lip}(\mathbb{R}^n)$  and for

any  $\psi \in C_{Lip}(\mathbb{R})$ , we define a sublinear expectation  $\mathbb{E}$  by

$$\mathbb{E}[\psi(\xi(\omega))] = \sup_{\{\sigma_i\} \in \Sigma(\underline{\sigma}, \bar{\sigma})} \mathbb{E}_{\mathbb{P}}[\psi(\varphi(\sigma_1 \epsilon_1, \dots, \sigma_n \epsilon_n))], \quad (\text{B1})$$

where  $\epsilon_1, \dots, \epsilon_n$  and  $\Sigma(\underline{\sigma}, \bar{\sigma})$  are as given in Problem 1 (and  $X_i = \sigma_i \epsilon_i$ .) It is clear that  $\mathbb{E}[\psi(X_1)] = \mathbb{E}[\psi(X_i)] = \sup_{\sigma_i \in [\underline{\sigma}, \bar{\sigma}]} \mathbb{E}[\psi(\sigma_i \epsilon_i)]$  for any Lipschitz  $\psi$ . Further, one can show that  $X_n$  is independent of  $X_1, \dots, X_{n-1}$  under  $\mathbb{E}$  using the definition of  $\Sigma(\underline{\sigma}, \bar{\sigma})$ . Finally, all the moment conditions are satisfied by the properties of classical normal distribution. Applying Peng's central limit theorem with (B1), we obtain Theorem 1.

## Acknowledgements

We thank the anonymous reviewers for their helpful comments.

## References

- Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- Zengjing Chen and Larry Epstein. Ambiguity, risk, and asset returns in continuous time. *Econometrica*, 70(4):1403–1443, 2002.
- Zengjing Chen, Tao Chen, Matt Davison, et al. Choquet expectation and Peng's g-expectation. *The Annals of Probability*, 33(3):1179–1199, 2005.
- Gustave Choquet. Theory of capacities. In *Annales de l'institut Fourier*, volume 5, pages 131–295, 1954.
- Laurent Denis, Mingshang Hu, and Shige Peng. Function spaces and capacity related to a sublinear expectation: application to g-brownian motion paths. *Potential Analysis*, 34(2):139–161, 2011.
- Larry G Epstein and Shaolin Ji. Ambiguous volatility, possibility and utility in continuous time. *Journal of Mathematical Economics*, 50:269–282, 2014.
- Xiao Fang, Shige Peng, Qi-Man Shao, and Yongsheng Song. Limit theorems with rate of convergence under sublinear expectations. *arXiv preprint arXiv:1711.10649*, 2017.
- Megan L Head, Luke Holman, Rob Lanfear, Andrew T Kahn, and Michael D Jennions. The extent and consequences of p-hacking in science. *PLoS Biology*, 13(3):e1002106, 2015.
- Shuo Huang and Gechun Liang. A monotone scheme for G-equations with application to the convergence rate of robust central limit theorem. *arXiv preprint arXiv:1904.07184*, 2019.
- Nicolai V Krylov. On Shige Peng's central limit theorem. *Stochastic Processes and their Applications*, 2019.
- Lu Lin, Yufeng Shi, Xin Wang, and Shuzhen Yang. k-sample upper expectation linear regression–modeling, identifiability, estimation and prediction. *Journal of Statistical Planning and Inference*, 170:15–26, 2016.
- Shige Peng. G-expectation, G-brownian motion and related stochastic calculus of Itô type. In *Stochastic Analysis and Applications*, pages 541–567. Springer, 2007.

- Shige Peng. A new central limit theorem under sublinear expectations. *arXiv preprint arXiv:0803.2656*, 2008.
- Shige Peng. Nonlinear expectations and stochastic calculus under uncertainty. *arXiv preprint arXiv:1002.4546*, 2010.
- Shige Peng. Law of large numbers and central limit theorem under nonlinear expectations. *Probability, Uncertainty and Quantitative Risk*, 4(1):4, 2019.
- Shige Peng, Shuzhen Yang, and Jianfeng Yao. Improving value-at-risk prediction under model uncertainty. *arXiv preprint arXiv:1805.03890*, 2018.
- Dmitry B Rokhlin. Central limit theorem under uncertain linear transformations. *Statistics & Probability Letters*, 107:191–198, 2015.
- Yongsheng Song. Normal approximation by Stein’s method under sublinear expectations. *Stochastic Processes and their Applications*, 2019.
- Jiongmin Yong and Xun Yu Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.