Statistics II **Problem Set 1**
Professor: Marcelo J. Moreira **Solutions**
TA: Luan Borelli **July 27, 2023**

---

1. *Show that $V[Y] = \mathbb{E}_X[V[Y|X]] + V_X[E[Y|X]]$.*

Solution. This is the Law of Total Variance. In what follows, the X subscript indicates that the expectation (or variance) operation is being taken with respect to the random variable X. When there is no subscript, the expected value is taken with respect to the joint distribution of X and Y.[1]

$$\begin{aligned}
V[Y] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] &= \mathbb{E}_X[\mathbb{E}[(Y - \mathbb{E}[Y|X] + \mathbb{E}[Y|X] - \mathbb{E}[Y])^2|X]] \\
&= \mathbb{E}_X[\mathbb{E}[(Y - \mathbb{E}[Y|X])^2|X] + 2\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - \mathbb{E}[Y])|X] + \mathbb{E}[(\mathbb{E}[Y|X] - \mathbb{E}[Y])^2|X]] \\
&= V_X[\mathbb{E}[Y|X]] + 2\mathbb{E}_X[\mathbb{E}[(Y - \mathbb{E}[Y|X])(\mathbb{E}[Y|X] - \mathbb{E}[Y])|X]] + \mathbb{E}_X[V[Y|X]] \\
&= V_X[\mathbb{E}[Y|X]] + 2\mathbb{E}_X[(\mathbb{E}[Y|X] - \mathbb{E}[Y])\underbrace{\mathbb{E}[(Y - \mathbb{E}[Y|X])|X]}_{0}] + \mathbb{E}_X[V[Y|X]] \\
&= V_X[\mathbb{E}[Y|X]] + \mathbb{E}_X[V[Y|X]].
\end{aligned}$$

Alternatively,

$$\begin{aligned}
\mathbb{E}_X[V[Y|X]] + V_X[\mathbb{E}[Y|X]] &= \mathbb{E}_X[\mathbb{E}_X[Y^2|X] - \mathbb{E}_X[Y|X]^2] + \mathbb{E}_X[\mathbb{E}_X[Y|X]^2] - \mathbb{E}_X[\mathbb{E}_X[Y|X]]^2 \\
&= \mathbb{E}_X[\mathbb{E}_X[Y^2|X]] - \mathbb{E}_X[\mathbb{E}_X[Y|X]^2] + \mathbb{E}_X[\mathbb{E}_X[Y|X]^2] - \mathbb{E}_X[\mathbb{E}_X[Y|X]]^2 \\
&= \mathbb{E}_X[\mathbb{E}_X[Y^2|X]] - \mathbb{E}_X[\mathbb{E}_X[Y|X]]^2 \\
\text{(L.I.E.)} &= \mathbb{E}_X[Y^2] - \mathbb{E}_X[Y]^2 = V[Y].
\end{aligned}$$

□

---

2. [7.1, LNs] *Let X_i be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $V[X_i] = \sigma^2$.*

(a) *Explain why $\bar{X}_N = N^{-1}\sum_{i=1}^N X_i$ converges in probability.*

Solution. Since $\mathbb{E}[\bar{X}_N] = \mu < \infty$, the weak law of large numbers applies. Thus $\bar{X}_N \xrightarrow{p} \mu$. If you consider it cheating to use the law of large numbers here, then take Chebyshev's inequality and note that, for any $\varepsilon > 0$,

$$0 \leq P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2/N}{\varepsilon^2} \to 0$$

as $N \to \infty$. Thus by definition of convergence in probability $\bar{X} \xrightarrow{p} \mu$. □

[1] That is, $\mathbb{E}_X[g(X,Y)] = \int_{-\infty}^{\infty} g(x,y) f_X(x)\, dx$ and $\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\, dx\, dy$.

(b) *Let $Z_N = N^{1/2}(\bar{X}_N - \mu)/\sigma$. Explain why $(Z_N)^2$ converges in distribution to a chi-square with one degree of freedom.*

Solution. Write $Z_n = \sigma^{-1}\sqrt{N}(\bar{X}_N - \mu)$. Since $\mathbb{E}[X] = \mu$ and $\mathbb{E}[X^2] = V[X] + \mathbb{E}[X]^2 = \sigma^2 + \mu^2 < \infty$, the central limit theorem applies and we have $\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{d} N(0,1)$. By Slutsky's theorem, it follows that $Z_n \xrightarrow{d} \sigma^{-1}N(0,\sigma^2) = N(0,1)$. Therefore, by continuity of the square function and the continuous mapping theorem $Z_n^2 \xrightarrow{d} Z^2$, where $Z \sim N(0,1)$. The result follows from the well-known fact that the square of a standard normally distributed random variable is chi-squared distributed with one degree of freedom; i.e., $Z^2 \sim \chi_1^2$. □

(c) *Let $W_N = N^{-1}\sum_{i=1}^N X_i^2$. Explain why W_N converges in distribution, and find its limiting distribution.*

Solution. As shown in **(b)**, $\mathbb{E}[X_i^2] = \sigma^2 + \mu^2 < \infty$. Moreover, since $\{X_i\}_{i=1}^N$ are i.i.d., then so are $\{X_i^2\}_{i=1}^N$.[2] Therefore, the strong law of large numbers applies and we have $W_N = N^{-1}\sum_{i=1}^N X_i^2 \xrightarrow{a.s.} \sigma^2 + \mu^2$. Since almost surely convergence implies convergence in distribution, it follows that $W_N \xrightarrow{d} \sigma^2 + \mu^2$. That is, W_N converges to a degenerate distribution: the point $\sigma^2 + \mu^2$. □

---

3. [7.4, LNs] *Let $Z_i = (X_i, Y_i)$ be i.i.d. random vectors where*

$$\mathbb{E}[Z_i] = \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix} \quad \textit{and} \quad V[Z_i] = \begin{bmatrix}\sigma_{XX} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{YY}\end{bmatrix}.$$

Assume for now that $\mu_Y \neq 0$.

(a) *Find the asymptotic distribution of $\sqrt{n}\left(\frac{\bar{X}_n}{\bar{Y}_n} - \frac{\mu_X}{\mu_Y}\right)$ using the multivariate delta method.*

Solution. Define $g : \mathbb{R} \times \mathbb{R}\backslash\{0\} \to \mathbb{R}$ as $g(x,y) = x/y$. Clearly g is twice continuously differentiable in $\mu = (\mu_X, \mu_Y)$, with gradient $\nabla g(\mu_X, \mu_Y) = (1/\mu_Y, -\mu_X/\mu_Y^2)'$. Since

$$\sqrt{n}\left(\begin{bmatrix}\bar{X}_n \\ \bar{Y}_n\end{bmatrix} - \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}\right) \xrightarrow{d} N\left(0, \begin{bmatrix}\sigma_{XX} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{YY}\end{bmatrix}\right),$$

it follows from the multivariate delta method that

$$\sqrt{n}\left(g(\bar{X},\bar{Y}) - g(\mu_X,\mu_Y)\right) \xrightarrow{d} \nabla g(\mu_X,\mu_Y)' N\left(0, \begin{bmatrix}\sigma_{XX} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{YY}\end{bmatrix}\right).$$

That is,

[2] If X_i, $i = 1,\ldots,N$, are i.i.d., then so are $h_i(X_i)$, for any measurable functions h_i, $i = 1,\ldots,N$.

$$\sqrt{n}\left(\frac{\bar{X}_n}{\bar{Y}_n}-\frac{\mu_X}{\mu_Y}\right)\xrightarrow{d} N\left(0,\begin{bmatrix}1/\mu_Y & -\mu_X/\mu_Y^2\end{bmatrix}\begin{bmatrix}\sigma_{XX} & \sigma_{XY}\\ \sigma_{XY} & \sigma_{YY}\end{bmatrix}\begin{bmatrix}1/\mu_Y\\ -\mu_X/\mu_Y^2\end{bmatrix}\right)$$
$$= N\left(0,\frac{1}{\mu_Y^4}\left(\sigma_{XX}\mu_Y^2-2\mu_X\mu_Y\sigma_{XY}+\sigma_{YY}\mu_X^2\right)\right).$$

□

(b) *Find the asymptotic distribution of* $\sqrt{n}\left(\frac{\bar{X}_n}{\mu_Y}-\frac{\mu_X}{\mu_Y}\right)=\sqrt{n}\left(\frac{\bar{X}_n-\mu_X}{\mu_Y}\right)$.

Solution. Write the above expression as $\mu_Y^{-1}\sqrt{n}(\bar{X}_n-\mu_X)$. By the central limit theorem, $\sqrt{n}(\bar{X}_n-\mu_X)\xrightarrow{d} N(0,\sigma_{XX})$. Therefore, by Slutsky's theorem,

$$\sqrt{n}\left(\frac{\bar{X}_n}{\mu_Y}-\frac{\mu_X}{\mu_Y}\right)=\mu_Y^{-1}\sqrt{n}(\bar{X}_n-\mu_X)\xrightarrow{d}\mu_Y^{-1}N(0,\sigma^2)=N(0,\mu_Y^{-2}\sigma_{XX}).$$

□

(c) *Why are the answers for items* ***(a)*** *and* ***(b)*** *in general different? Explain why they are the same when* $\mu_X=0$.

Solution. In **(b)** we replaced the sample mean $\bar{Y}_n$, which is a consistent estimator for the population mean, by the population mean itself. $\bar{Y}_n$, being an estimator, possesses some variance that influences the asymptotic variance of $\sqrt{n}\left(\frac{\bar{X}_n}{\bar{Y}_n}-\frac{\mu_X}{\mu_Y}\right)$. When replaced by the true parameter μ_Y, all such variance vanishes and we obtain the asymptotic variance observed in **(b)**. Notice that if we set $\sigma_{XY}=\sigma_{YY}=0$ in **(a)** we obtain the asymptotic variance of **(b)**.

When $\mu_X=0$ the asymptotic variance of $\sqrt{n}\left(\frac{\bar{X}_n}{\bar{Y}_n}-\frac{\mu_X}{\mu_Y}\right)$ does not depend on σ_{XY} and σ_{YY}, so the variance reduction obtained by replacing $\bar{Y}_n$ by μ_Y does not influence the asymptotic variance anymore and the asymptotic variances obtained in **(a)** and **(b)** become equivalent. The independence of the asymptotic variance of $\sqrt{n}\left(\frac{\bar{X}_n}{\bar{Y}_n}-\frac{\mu_X}{\mu_Y}\right)$ with respect to σ_{XY} and σ_{YY} when $\mu_X=0$ reflects the fact that $\bar{Y}_N$ interacts scalarly with $\bar{X}_n$. The intuition is that as $n\to\infty$, $\bar{X}_n$ converges in probability to zero, making the "scalar effects" of $\bar{Y}_n$ on $\bar{X}_n$ disappear. □

(d) *What happens with* $\sqrt{n}\bar{X}_n/\bar{Y}_n$ *when* $\mu_X=\mu_Y=0$*? How about the asymptotic distribution of* $\bar{X}_n/\bar{Y}_n$*? Explain your answer.*

Solution. The function g defined in **(a)** is not continuously differentiable at $(\mu_X,\mu_Y)=(0,0)$. Therefore we cannot apply the (multivariate) delta method. Regarding the asymptotic distribution of $\bar{X}_n/\bar{Y}_n$, when $(\mu_X,\mu_Y)=(0,0)$, by the central limit theorem we have that

$$\sqrt{n}\bar{X}_n\xrightarrow{d} N(0,\sigma_{XX})\quad\text{and}\quad\sqrt{n}\bar{Y}_n\xrightarrow{d} N(0,\sigma_{YY}).$$

Using the same g as defined in **(a)**, which is clearly continuous almost everywhere, by the continuous mapping theorem we have

$$g(\sqrt{n}\bar{X}_n, \sqrt{n}\bar{Y}_n) = \frac{\sqrt{n}\bar{X}_n}{\sqrt{n}\bar{Y}_n} = \frac{\bar{X}_n}{\bar{Y}_n} \xrightarrow{d} \frac{Z_X}{Z_Y},$$

where $Z_X \sim N(0, \sigma_{XX})$ and $Z_Y \sim N(0, \sigma_{YY})$.

What is the distribution of Z_X/Z_Y? It turns out that it is a Cauchy distribution. The Cauchy distribution does not have finite moments of order greater than or equal to one (only fractional absolute moments exist) and has no moment generating function! □

---

4. *Prove Result 10.3 from the lecture notes.*

Solution. See lecture notes, p. 62. □

---

5. [12.1, LNs] *Suppose that $X_1, \dots, X_n$ are i.i.d., each uniformly distributed on $[0, \theta]$.*

(a) *Find the asymptotic distribution of $n^{1/2}(\theta - U_n)$, where $U_n = 2(X_1 + \cdots + X_n)/n$. Hint: find the mean and variance of X_i.*

Solution. Notice that

$$n^{1/2}(\theta - U_n) = n^{1/2}\left(\theta - \frac{2}{n}\sum_{i=1}^{n} X_i\right) = -2n^{1/2}\left(\bar{X} - \frac{\theta}{2}\right).$$

Since $\mathbb{E}[X_i] = \frac{\theta}{2}$ and $\mathbb{E}[X_i^2] = V[X_i] + \mathbb{E}[X_i]^2 = \frac{\theta^2}{12} + \frac{\theta^2}{4} < \infty$, CLT applies. Therefore

$$n^{1/2}(\theta - U_n) \xrightarrow{d} -2N\left(0, \frac{\theta^2}{12}\right) = N\left(0, \frac{\theta^2}{3}\right),$$

by the central limit theorem, together with Slutsky's theorem. □

(b) *Find the (minimally) sufficient statistic for θ.*

Solution. The joint density of $X \equiv X_1, \dots, X_n$ is $f_X(X;\theta) = \prod_{i=1}^n \frac{1}{\theta} = \frac{1}{\theta^n}$ if $0 < X_i < \theta$ for all i and zero otherwise. This can be compactly written as

$$f_X(X;\theta) = \frac{1}{\theta^n}\mathbf{1}\{0 < \min(X_1, \dots, X_n)\}\mathbf{1}\{\max(X_1, \dots, X_n) < \theta\}.$$

Therefore by Fisher's factorization theorem $T(X) \equiv \max(X_1, \dots, X_n)$ is a sufficient statistic for θ. Moreover, it is a minimal sufficient statistic. Indeed, for any X and Y

$$\frac{f(X;\theta)}{f(Y;\theta)} = \frac{\mathbf{1}\{0 < \min(X_1, \dots, X_n)\}\mathbf{1}\{\max(X_1, \dots, X_n) < \theta\}}{\mathbf{1}\{0 < \min(Y_1, \dots, Y_n)\}\mathbf{1}\{\max(Y_1, \dots, Y_n) < \theta\}}$$

is constant as a function of θ if and only if $\max(X_1, \ldots, X_n) = \max(Y_1, \ldots, Y_n)$. Therefore, by Lehmann and Scheffé (1950, Theorem 6.3) $T(X)$ is a minimal sufficient statistic.[3] □

(c) *Find the mean and variance of* $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

Solution. The cumulative distribution function of $X_{(n)}$ is given by

$$\begin{aligned} F(x;\theta) \equiv P(\max\{X_1, \ldots, X_n\} \leq x) &= P(X_1 \leq x, \ldots, X_n \leq x) \\ \text{(Independence)} &= \prod_{i=1}^{n} P(X_i \leq x) \\ \text{(Identicality)} &= \prod_{i=1}^{n} \frac{x}{\theta} = \left(\frac{x}{\theta}\right)^n \end{aligned}$$

for $x \in [0, \theta]$ and zero otherwise. The probability density function is then given by

$$f(x;\theta) = n\frac{x^{n-1}}{\theta^n} \quad \text{for } x \in [0,\theta] \quad \text{and zero otherwise.}$$

Therefore the mean is given by

$$\mathbb{E}[X_{(n)}] = \int_0^\theta xn\frac{x^{n-1}}{\theta^n}\, dx = \frac{n}{n+1}\theta.$$

For the variance, notice that

$$\mathbb{E}[X_{(n)}^2] = \int_0^\theta x^2 n\frac{x^{n-1}}{\theta^n}\, dx = \frac{n}{n+2}\theta^2,$$

whence it follows that

$$V[X_{(n)}] = \mathbb{E}[X_{(n)}^2] - \mathbb{E}[X_{(n)}]^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2\theta^2.$$

□

(d) *Show that* $n^{1/2}(\theta - X_{(n)}) \xrightarrow{p} 0$.

Solution. Notice that

$$\begin{aligned} P(|n^{1/2}(\theta - X_{(n)})| \geq \varepsilon) &= P(\sqrt{n}(\theta - X_{(n)}) \leq \theta - \frac{\varepsilon}{n^{1/2}}) \\ \text{(CDF)} &= \frac{\left(\theta - \varepsilon/n^{1/2}\right)^n}{\theta^n} = \left(1 - \frac{\varepsilon/\theta}{n^{1/2}}\right)^n = \left(\left(1 - \frac{\varepsilon/\theta}{n^{1/2}}\right)^{\sqrt{n}}\right)^{\sqrt{n}}. \end{aligned}$$

As $n \to \infty$, $\left(1 - \frac{\varepsilon/\theta}{n^{1/2}}\right)^{\sqrt{n}} \to \exp(-\varepsilon/\theta)$ and hence $P(|n^{1/2}(\theta - X_{(n)}| \geq \varepsilon) \to 0$. Therefore $n^{1/2}(\theta - X_{(n)}) \xrightarrow{p} 0$.[4] □

[3] For Lehmann and Scheffé's theorem, I refer to Casella and Berger's book. See p. 281, Theorem 6.2.13.

[4] If you are not convinced by this heuristic argument, write

$$\lim_{n\to\infty}\left(1 - \frac{\varepsilon/\theta}{\sqrt{n}}\right)^n = \lim_{n\to\infty}\exp\left(n\ln\left(1 - \frac{\varepsilon/\theta}{\sqrt{n}}\right)\right) = \exp\left(\lim_{n\to\infty}\ln\left(1 - \frac{\varepsilon/\theta}{\sqrt{n}}\right)/n^{-1}\right),$$

and then apply L'Hôpital rule twice. I left these boring calculations to you.

(e) *Find the asymptotic distribution of $n(\theta - X_{(n)})$. Hint: it may be convenient to work directly with the cdf of $n(\theta - X_{(n)})$.*

Solution. We have that

$$\begin{aligned} P(n(\theta - X_{(n)}) \le x) &= P(X_{(n)} \ge x/n - \theta) \\ &= 1 - P(X_{(n)} \le \theta - x/n) \\ &= 1 - \frac{(\theta - x/n)^n}{\theta^n} \\ &= 1 - \left(1 - \frac{x/n}{n}\right)^n \to 1 - \exp(-x/\theta) \end{aligned}$$

as $n \to \infty$, which is the CDF of an exponential distribution with parameter $1/\theta$. □

(f) *Compare the asymptotic distribution of U_n and $X_{(n)}$, derived respectively in items **(a)** and **(b)**.*

Solution. U_n is (asymptotically) normally distributed. Although $n(\theta - X_n)$ is (asymptotically) exponentially distributed, from this result, there is not much that can be said about the (asymptotic) distribution of $X_{(n)}$. I confess I don't know the exact answer, but I suspect it has something to do with extremum value distributions. Any suggestion is welcome. □

---

6. [14.7, LNs] *Assume that $X_i \overset{\text{iid}}{\sim} N(\theta, 1)$.*

(a) *Show that the joint pdf is a special case of an exponential family.*

Solution. The single-parameter exponential family is the class of probability distributions whose probability density function (or probability mass function, in case of discrete distributions) can be expressed as

$$f(X; \theta) = C(\theta) \exp(A(\theta) T(X)) h(X),$$

where $C(\theta)$, $A(\theta)$ and $h(X)$ are known functions and $C(\theta)$ is non-negative. Since X_i, for $i = 1, \ldots, n$, are i.i.d., the joint pdf of $X \equiv (X_1, \ldots, X_n)$ is given by

$$\begin{aligned} f(X; \theta) &= \prod_{i=1}^{n} (2\pi)^{-1/2} \exp\left(-\frac{(X_i - \theta)^2}{2}\right) \\ &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i^2 - 2\theta X_i + \theta^2)\right) \\ &= (2\pi)^{-n/2} \exp\left(-\frac{n}{2}\theta\right) \exp\left(\theta \sum_{i=1}^{n} X_i\right) \exp\left(-\frac{1}{2} \sum_{i=1}^{n} X_i^2\right). \end{aligned}$$

By setting

$$C(\theta) \equiv \exp\left(\theta \sum_{i=1}^{n} X_i\right), \quad A(\theta) \equiv \theta, \quad T(X) \equiv \sum_{i=1}^{n} X_i, \quad h(X) \equiv (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} X_i^2\right)$$

it becomes clear that this joint pdf is a special case of the exponential family. □

(b) *Show that the UMP test for $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ reject the null when $\sqrt{n}(\bar{X}_n - \theta_0) > c_{1-\alpha}$, where $c_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal.*

Solution. For distributions of the exponential family with $A(\theta)$ monotone increasing, there exists a UMP for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ characterized by the critical region

$$C_X = \{X \mid T(X) > k\},$$

where k is determined by $\alpha = \int_{C_X} f(X; \theta_0)\, dx$.[5] In particular, from **(a)** we have $A(\theta) = \theta$, which is clearly monotone increasing, and $T(X) = \sum_{i=1}^{n} X_i$. Therefore the UMP test is characterized by the critical region

$$\begin{aligned} C_X &= \left\{ X \;\middle|\; \sum_{i=1}^{n} X_i > k \right\} \\ &= \left\{ X \;\middle|\; \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \theta_0\right) > k' \right\}. \end{aligned}$$

Under the null, $\bar{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\theta_0, \frac{1}{n})$ and hence $\sqrt{n}(\bar{X} - \theta_0) \sim N(0,1)$. It follows that

$$\int_{C_X} f(X; \theta_0)\, dx = 1 - \Phi(k') = \alpha \iff k' = \Phi^{-1}(1-\alpha),$$

which is the $1-\alpha$ quantile of a standard normal distribution. □

(c) *Suppose someone discards the even observations and constructs a one-sided test using averages of the odd observations. Compare the asymptotic power (using Pitman's drift) of this test with the test using averages of all observations found in part* ***(b)****.*

Solution. Define the sequence $\theta_n = \theta_0 + \frac{h}{\sqrt{n}}$ of local alternatives. For the test using averages of all observations we have

$$\begin{aligned} \sqrt{n}(\bar{X} - \theta_0) &= \sqrt{n}(\bar{X} - \theta_n + \theta_n - \theta_0) \\ &= \sqrt{n}(\bar{X} - \theta_n) + \sqrt{n}(\theta_n - \theta_0) \\ &= \sqrt{n}(\bar{X} - \theta_n) + h. \end{aligned}$$

[5] For further details I refer to Lehmann and Romano's book "Testing Statistical Hypothesis". See Theorem 3.4.1 and, more specifically, Corollary 3.4.1. A rigorous reader will note that the test of Corollary 3.4.1 is actually a randomized test that also involves rejecting the null hypothesis with some probability γ when $T(x) = k$. Note, however, that for the particular case of this question, $T(X)$ is continuous when viewed as a random variable. Therefore, $T(X) = k$ implies a set of measure zero.

Notice that as $n \to \infty$, $\theta_n \to \theta_0$ and hence $\sqrt{n}(\bar{X} - \theta_n) \xrightarrow{d} N(0,1)$ by the central limit theorem. Therefore$\sqrt{n}(\bar{X} - \theta_0) \xrightarrow{d} Z + h \sim N(h,1)$, where $Z \sim N(0,1)$. The asymptotic local power is given by

$$\begin{aligned}\lim_{n\to\infty} P(\sqrt{n}(\bar{X} - \theta_0) \geq c) &= P(Z + h \geq c) = P(Z \geq c - h) \\ &= 1 - P(Z < c - h) = 1 - \Phi(c - h) \\ &= \Phi(h - c).\end{aligned}$$

For the one-sided test using averages of the "odd" observations only, suppose, without loss of generality, that n is even. By discarding the "even" observations we obtain the new statistic

$$\bar{X}' = \frac{1}{(n/2)} \sum_{i=1}^{n/2} X_i.$$

By proceeding in the same way as before, we obtain

$$\sqrt{n}(\bar{X}' - \theta_0) = \sqrt{n}(\bar{X}' - \theta_n) + h.$$

Now $\mathbb{E}[\bar{X}'] = \theta$ and $V[\bar{X}'] = 2/n$. Thus, as $n \to \infty$, $\sqrt{n}(\bar{X} - \theta_n) \xrightarrow{d} N(0,2)$. Therefore $\sqrt{n}(\bar{X}' - \theta_0) \xrightarrow{d} \sqrt{2}Z + h \sim N(h,2)$. The asymptotic local power then becomes

$$\begin{aligned}\lim_{n\to\infty} P(\sqrt{n}(\bar{X}' - \theta_0) \geq c) &= P(\sqrt{2}Z + h \geq c) = P\left(Z \geq \frac{c-h}{\sqrt{2}}\right) \\ &= 1 - P\left(Z < \frac{c-h}{\sqrt{2}}\right) = 1 - \Phi\left(\frac{c-h}{\sqrt{2}}\right) \\ &= \Phi\left(\frac{h-c}{\sqrt{2}}\right).\end{aligned}$$

Denote $\delta \equiv h - c$. Since $\sqrt{2} > 1$ and Φ is strictly increasing, it follows that

$$\Phi(\delta) > \Phi\left(\frac{\delta}{\sqrt{2}}\right) \quad \forall \delta \in \mathbb{R}.$$

That is, the test based on the full sample statistic $\bar{X}$ is (asymptotically) uniformly more powerful than the test based on the half-sample statistic $\bar{X}'$. □

(d) *Show that the UMP test for $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$ reject the null when $\sqrt{n}(\bar{X}_n - \theta_0) < c_\alpha$. How is c_α related to $c_{1-\alpha}$.*

Solution. In this case, the UMP is characterized by the critical region

$$\begin{aligned}C_X &= \left\{ X \;\middle|\; \sum_{i=1}^{n} X_i < k \right\} \\ &= \left\{ X \;\middle|\; \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \theta_0\right) < k' \right\}.\end{aligned}$$

Therefore

$$\int_{C_X} f(X;\theta_0)\ dx = \Phi(k') = \alpha \iff k' = \Phi^{-1}(\alpha),$$

which is the α quantile of a standard normal distribution.

Observe that if c_α is the α quantile and $c_{1-\alpha}$ the $1-\alpha$ quantile of a standard normal distribution, then $\Phi(c_\alpha) = \alpha$, $\Phi(c_{1-\alpha}) = 1-\alpha$ and hence $\Phi(c_{1-\alpha}) = 1 - \Phi(c_\alpha) = \Phi(-c_\alpha)$. The last equality follows from the symmetry of the standard normal distribution. Therefore $c_{1-\alpha} = -c_\alpha$. □

7. [15.5, LNs] *Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution on the positive numbers with pdf $f(x,\theta) = \theta^2 x \exp(-\theta x)$, with $\theta > 0$.*

(a) *Find the maximum likelihood estimator (MLE) $\hat{\theta}_n$ for θ.*

Solution. The log-likelihood function is given by

$$\begin{aligned} L_n(X;\theta) &= \ln \prod_{i=1}^{n} \theta^2 X_i \exp(-\theta X_i) = \sum_{i=1}^{n} \ln(\theta^2 X_i \exp(-\theta X_i)) \\ &= \sum_{i=1}^{n} [\ln\theta^2 + \ln X_i - \theta X_i] = n \ln\theta^2 + \sum_{i=1}^{n} \ln X_i - \theta \sum_{i=1}^{n} X_i. \end{aligned}$$

The first-order condition for the problem of maximizing $L_n(X;\theta)$ with respect to θ is

$$\frac{dL_n(X;\theta)}{d\theta} = n\frac{2}{\theta} - \sum_{i=1}^{n} X_i = 0,$$

whence it follows that the maximum likelihood estimator is

$$\hat{\theta}_n = \frac{2}{\frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{2}{\bar{X}}.$$

The second-order condition will be satisfied. Believe me. □

(b) *Is the MLE $\hat{\theta}$ the minimum variance unbiased estimator (MVUE) for θ? Explain your answer.*

Solution. No. For $\hat{\theta}_n$ to be MVUE it must be unbiased, which is not the case. Observe that

$$\mathbb{E}[\bar{X}] = \mathbb{E}[X_i] = \int_0^\infty \theta^2 x^2 \exp(-\theta x)\ dx = \frac{2}{\theta}.$$

Define $g : \mathbb{R}_{++} \to \mathbb{R}$ as $g(x) = 2/x$. This function is strictly convex. Therefore, by Jensen's inequality,

$$\mathbb{E}[\hat{\theta}_n] = \mathbb{E}[2/\bar{X}] = \mathbb{E}[g(\bar{X})] > g(\mathbb{E}[\bar{X}]) = \frac{2}{(2/\theta)} = \theta.$$

□

(c) *Recall that* $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I(\theta_0)^{-1})$. *Find the scalar* $I(\theta_0)$.

Solution. By definition, $I(\theta_0) = -\mathbb{E}\left[\frac{\partial^2 L(\theta_0)}{\partial \theta_0^2}\right] = \mathbb{E}\left[-2/\theta_0^2\right] = 2/\theta_0^2$, where L is the single observation log-likelihood function. □

(d) *Construct a 95% confidence region using the Wald test which rejects the null when* $n \cdot I(\hat{\theta}_n)(\hat{\theta}_n - \theta_0)^2 > c_\alpha$.

Solution. We have that $I(\hat{\theta}_n) = 2/\hat{\theta}_n^2$, so the test statistic is given by

$$n \cdot I(\hat{\theta}_n)(\hat{\theta}_n - \theta_0)^2 = n\frac{2}{\hat{\theta}_n^2}(\hat{\theta}_n - \theta_0)^2 = 2n\left(1 - \frac{\theta_0}{\hat{\theta}_n}\right)^2.$$

We reject the null if

$$\begin{aligned} 2n\left(1 - \frac{\theta_0}{\hat{\theta}_n}\right)^2 > c_\alpha &\iff \left|1 - \frac{\theta_0}{\hat{\theta}_n}\right| > \sqrt{\frac{c_\alpha}{2n}} \\ \iff \frac{\theta_0}{\hat{\theta}_n} > 1 + \sqrt{\frac{2c_\alpha}{n}} &\quad \text{or} \quad \frac{\theta_0}{\hat{\theta}_n} < 1 - \sqrt{\frac{2c_\alpha}{n}} \\ \iff \theta_0 > \left(1 + \sqrt{\frac{2c_\alpha}{n}}\right)\hat{\theta}_n &\quad \text{or} \quad \theta_0 < \left(1 - \sqrt{\frac{2c_\alpha}{n}}\right)\hat{\theta}_n. \end{aligned}$$

Therefore the critical region is

$$C_X = \left\{X \;\middle|\; \theta_0 > \left(1 + \sqrt{\frac{2c_\alpha}{n}}\right)\hat{\theta}_n \quad \text{or} \quad \theta_0 < \left(1 - \sqrt{\frac{2c_\alpha}{n}}\right)\hat{\theta}_n\right\},$$

and hence the confidence region

$$\begin{aligned} C_X^C &= \left\{X \;\middle|\; \left(1 - \sqrt{\frac{2c_\alpha}{n}}\right)\hat{\theta}_n \le \theta_0 \le \left(1 + \sqrt{\frac{2c_\alpha}{n}}\right)\hat{\theta}_n\right\} \\ &= \left\{X \;\middle|\; \left(1 - \sqrt{\frac{2c_\alpha}{n}}\right)\frac{2}{\bar{X}} \le \theta_0 \le \left(1 + \sqrt{\frac{2c_\alpha}{n}}\right)\frac{2}{\bar{X}}\right\}. \end{aligned}$$

For a 95% confidence region, just set $c_\alpha = 3.84$. □

(e) *Construct a 95% confidence region using the Wald test which rejects the null when* $n \cdot I(\theta_0)(\hat{\theta}_n - \theta_0)^2 > c_\alpha$.

Solution. We have that $I(\theta_0) = 2/\theta_0^2$, so the test statistic is given by

$$n\frac{2}{\theta_0^2}(\hat{\theta}_n - \theta_0)^2 = 2n\left(\frac{\hat{\theta}_n}{\theta_0} - 1\right)^2.$$

We reject the null if

$$
\begin{aligned}
2n\left(\frac{\hat{\theta}_n}{\theta_0}-1\right)^2 > c_\alpha \iff \left|\frac{\hat{\theta}_n}{\theta_0}-1\right| > \sqrt{\frac{c_\alpha}{2n}} \\
\iff \frac{\hat{\theta}_n}{\theta_0} > 1+\sqrt{\frac{c_\alpha}{2n}} \quad \text{or} \quad \frac{\hat{\theta}_n}{\theta_0} < 1-\sqrt{\frac{c_\alpha}{2n}} \\
\iff \frac{\hat{\theta}_n}{\theta_0} > 1+\sqrt{\frac{c_\alpha}{2n}} \quad \text{or} \quad \frac{\hat{\theta}_n}{\theta_0} < 1-\sqrt{\frac{c_\alpha}{2n}} \\
\iff \theta_0 < \left(1+\sqrt{\frac{c_\alpha}{2n}}\right)^{-1}\hat{\theta}_n \quad \text{or} \quad \theta_0 > \left(1-\sqrt{\frac{c_\alpha}{2n}}\right)^{-1}\hat{\theta}_n.
\end{aligned}
$$

Therefore the critical region is

$$
C_X = \left\{ X \;\middle|\; \theta_0 < \left(1+\sqrt{\frac{c_\alpha}{2n}}\right)^{-1}\hat{\theta}_n \quad \text{or} \quad \theta_0 > \left(1-\sqrt{\frac{c_\alpha}{2n}}\right)^{-1}\hat{\theta}_n \right\},
$$

and hence the confidence region

$$
\begin{aligned}
C_X &= \left\{ X \;\middle|\; \left(1+\sqrt{\frac{c_\alpha}{2n}}\right)^{-1}\hat{\theta}_n \le \theta_0 \le \left(1-\sqrt{\frac{c_\alpha}{2n}}\right)^{-1}\hat{\theta}_n \right\} \\
&= \left\{ X \;\middle|\; \left(1+\sqrt{\frac{c_\alpha}{2n}}\right)^{-1}\frac{2}{\bar{X}} \le \theta_0 \le \left(1-\sqrt{\frac{c_\alpha}{2n}}\right)^{-1}\frac{2}{\bar{X}} \right\}.
\end{aligned}
$$

For a 95% confidence region, just set $c_\alpha = 3.84$. □

(f) *How is the MLE $\hat{\phi}_n$ for $\phi = \ln\theta$ related to $\hat{\theta}_n$?*

Solution. Maximum likelihood estimators satisfy the invariance property: for any function τ, if $\hat{\theta}_n$ is the maximum likelihood estimator of θ, then $\tau(\hat{\theta}_n)$ is the maximum likelihood estimator of $\tau(\theta)$. Therefore $\hat{\phi}_n = \ln\hat{\theta}_n$. □

8. [7.9, LNs] *Let $X \sim N(\theta, 1)$. The density of X is $g(x;\theta) = (2\pi)^{-1/2} \exp\left(-(x-\theta)^2/2\right)$.*

(a) *Show that the density ratio $g(X;\theta)/h(x)$ has mean 1 when X is being drawn from the density $h(x)$.*

Solution.

$$\mathbb{E}\left[\frac{g(X;\theta)}{h(x)}\right] = \int_{-\infty}^{\infty} \frac{g(x;\theta)}{h(x)} h(x)\, dx = \int_{-\infty}^{\infty} g(x;\theta)\, dx = 1.$$

□

(b) *Let $X_i \overset{\text{iid}}{\sim} N(0,1)$. Report on a table the sample average you found of $n^{-1} \sum_{i=1}^{n} \frac{g(X_i;\theta)}{g(X_i;0)}$ for all combinations of $n = 10^j$ for $j = 1,2,3,4,5,6$ and $\theta = 0,1,2,3,4,5,6,7,8,9$. Here, $h(x)$ is the density of a standard normal.*

n/θ	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
10	1	1.5284	0.4357	0.3125	0.0001	0.0536	0.000010	0.000000008	0.000000000022	0.000000002230
100	1	0.9669	0.9024	0.5465	0.0345	0.0916	0.000478	0.000056420	0.000000021593	0.000000002721
1000	1	0.9739	1.3913	0.7259	0.4612	0.1379	0.005234	0.000055074	0.000000429973	0.000000000436
10000	1	0.9908	1.1457	0.7672	0.3040	0.4634	1.412345	0.000375722	0.000026128897	0.000000022828
100000	1	1.0009	0.9915	0.9957	1.1448	1.7508	0.273413	0.060356958	0.000537437971	0.000016813355
1000000	1	1.0009	0.9915	0.9957	1.1448	1.7508	0.273413	0.0603569584	0.0005374379711	0.0000168133549

(c) *Explain your results by plotting densities of $N(0,1)$, $N(3,1)$, $N(6,1)$, and $N(9,1)$.*

Solution. The goal of this question is to estimate the mean of a density ratio of the form $g(X;\theta)/h(X)$, where $X \sim N(0,1)$. From item **(a)**, we know that the true population mean of this density ratio is always 1, irrespective of the specific density function $h(X)$. Thus, according to the law of large numbers, we would expect the sample means from the results in table **(b)** to converge to 1 as the sample size n increases, regardless of the value of θ. Indeed, this convergence is observed in some cases, particularly for small values of θ (e.g., from 1 to 4). However, for larger values of θ, the convergence seems to break down. For instance, when $\theta = 9$, even with a very large sample size like $n = 1,000,000$, the estimated mean remains *extremely* distant from 1.[6]

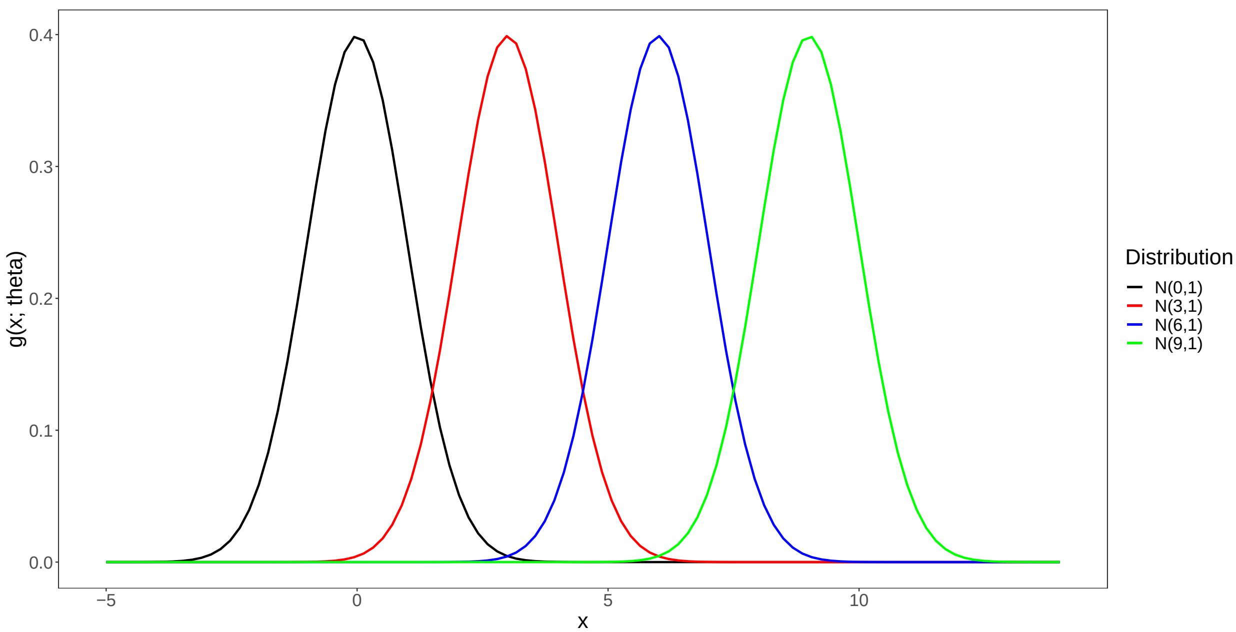
What's happening? Is the law of large numbers failing for very high values of θ? Figure 1 can help us address this question. By looking at the density of $N(0,1)$ and comparing it to the density of $N(9,1)$, and considering the ratio $g(X_i;9)/g(X_i;0)$ as an example, we can observe that since $X_i \sim N(0,1)$, on average, the draws will be close to zero. Thus, for the vast majority of draws, $g(X_i;0)$ will take a high value, while $g(X_i;9)$ will take an extremely low value. As a result, in the vast majority of draws, the ratio $g(X_i;9)/g(X_i;0)$ will assume a very low value, virtually zero. However, in the occurrence of extremely rare events where positive draws of X_i deviate very far from zero, the logic will be reversed: $g(X_i;9)$ will take a high value, and $g(X_i;0)$ will take an extremely low value, causing the ratio to skyrocket to an *extremely* high value. This outlier will force the sample mean upwards, bringing it again

[6] I even tested with an even larger sample size of $n = 50,000,000$, but there was hardly any difference. The estimate remains significantly far from 1.

closer to 1 and compensating for all the previous observations.

Now it becomes clear what is actually occurring: it is not that the law of large numbers fails for high values of θ; rather, the issue lies in the fact that, for high values of θ, the realizations of "crucial importance" for the convergence of the sample mean to the population mean are *extremely rare*. The convergence happens due to an exceedingly small number of extremely rare events, which carry a disproportionately significant impact compared to the rest of the events. As a consequence, in order to computationally observe the convergence for high values of θ, an enormously large number of observations would be required — practically approaching infinity! This is necessary to ensure that these exceptionally rare events happen frequently enough to "drive the convergence." However, from a computational standpoint, performing such an immense number of simulations can be impractical. □

Figure 1: Normal densities $g(x;\theta)$ with $\theta = 0, 3, 6$ and 9.



(d) *Show that* $h(x) = 10^{-1}\sum_{j=0}^{9} g(x;j)$ *is a density. Furthermore, how would you draw a random sample from* $h(x)$ *in practice?*

Solution. For the sake of simplicity, let's abstract from technicalities involving the formal definition of a density function by just assuming that for $h(x)$ to be a density it suffices to show that (i) $h(x)$ integrates to 1 over $(-\infty, \infty)$ and (ii) $h(x)$ is nonnegative for all x.

We have that

$$\int_{-\infty}^{\infty} 10^{-1} \sum_{j=0}^{9} g(x;j)\, dx = \int_{-\infty}^{\infty} 10^{-1} \sum_{j=0}^{9} \exp(-(x-\theta)^2/2)\, dx \tag{1}$$

$$= 10^{-1} \sum_{j=0}^{9} \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp(-(x-\theta)^2/2)\, dx \tag{2}$$

$$= 10^{-1} \sum_{j=0}^{9} 1 = 10^{-1} 10 = 1. \tag{3}$$

Moreover, since g is a density, $g(x;j) \geq 0$ for all x and $j = 0, \ldots, 9$, whence it follows that $h(x) \geq 0$ for all x. Therefore, h is a density.

In order to draw a random sample from $h(x)$ we could use inverse transform sampling. In practice, we simply generate a draw u from a *discrete* uniform distribution in the interval $[0, 9]$ and then generate a draw from a normal distribution with mean u and variance 1; that is, a draw from $N(u, 1)$. The resulting draw will be equivalent to a draw from $h(x)$. □

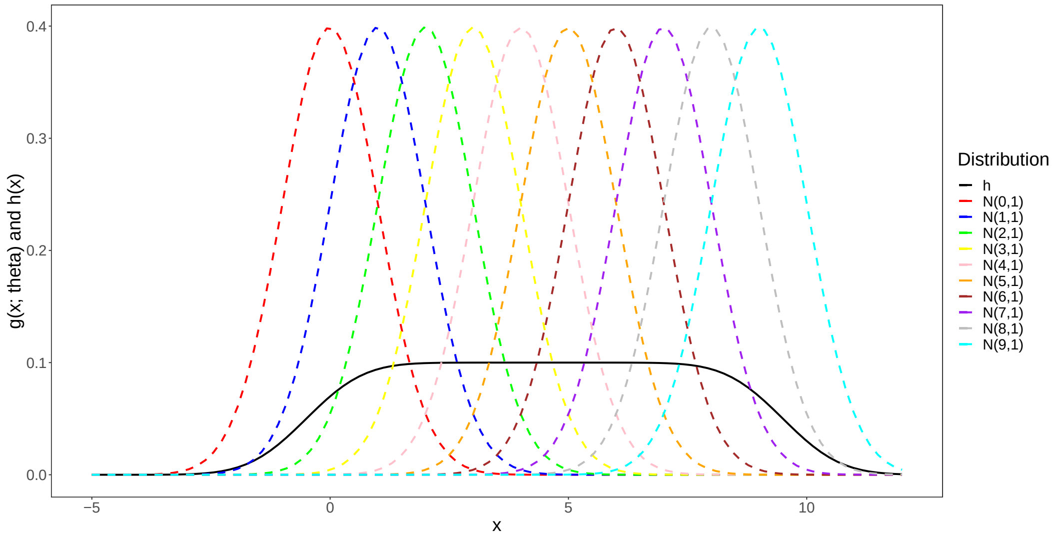
(e) *Report on a table the sample average you found of* $n^{-1} \sum_{i=1}^{n} \frac{g(X_i;\theta)}{h(X_i)}$ *for all combinations of* $n = 10^j$ *for* $j = 1, 2, 3, 4, 5, 6$ *and* $\theta = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$. *Here* $h(x) = 10^{-1} \sum_{j=0}^{9} g(x;j)$.

n/θ	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
10	0.713267	0.876365	0.619611	0.848035	0.998121	0.769224	0.751788	0.54929	0.641165	1.114872
100	1.068189	1.045083	1.052162	1.248939	1.022221	0.847806	0.97673	0.827878	1.182188	0.799998
1000	1.108903	0.994146	0.934179	0.980026	0.918329	1.070847	1.023598	1.035077	1.02438	0.987447
10000	0.99591	1.000614	1.000128	0.988394	0.984994	1.002482	0.983719	1.002187	1.003022	1.001579
100000	1.009509	1.001897	1.007725	0.999622	1.001571	0.996998	1.005517	0.999736	1.004651	1.003114
1000000	0.999541	0.998239	1.000453	1.000378	1.000569	0.998407	0.999694	0.999728	0.998736	0.99614

(f) *Explain your results by plotting the densities of* $h(x)$ *and of* $N(0,1)$, $N(3,1)$, $N(6,1)$, *and* $N(9,1)$.

Solution. One way to address the problem described in item **(c)** is to generate draws from a distribution that assigns greater "importance" (i.e., higher frequency) to the values that "truly matter" for determining the convergence of the sample mean of the density ratio. The distribution defined in item **(d)** accomplishes precisely that. Continuing with the example of the case $\theta = 9$ described in item **(c)**, when the draws are generated from $h(x)$ instead of $g(x;0)$, higher values of the numerator $g(X_i;9)$ occur more frequently, and the convergence of the sample mean to 1 happens without relying as much on the occurrence of extremely rare events. Figure 2 illustrates how the distribution $h(x)$ "prioritizes" the sampling in more important regions of $g(x;9)$ for determining the convergence when compared to $g(x;0)$. This method has a name: importance sampling. □

Figure 2: Normal densities $g(x;\theta)$ with $\theta = 1, \ldots, 9$ and mixture density $h(x)$.



9. [2.5, Hansen] *Show that $\sigma^2(X)$ is the best predictor of e^2 given X:*

(a) *Write down the mean-squared error of a predictor $h(X)$ for e^2.*

Solution. $\mathbb{E}[(e^2 - h(X))^2]$. □

(b) *What does it mean to be predicting e^2?*

Solution. The CEF error is defined as $e = Y - m(X)$, where $m(X) = \mathbb{E}[Y|X]$. Therefore, to be predicting e^2 means to be predicting $(Y - \mathbb{E}[Y|X])^2$, which can be understood as the squared prediction error of Y when considering the CEF as a predictor of Y. Recall that the CEF is the best predictor of Y. Therefore, to be predicting e^2 means to be predicting the squared prediction error of Y when considering the best possible predictor of Y. □

(c) *Show that $\sigma^2(X)$ minimizes the mean-squared error and is thus the best predictor.*

Solution. Note that

$$\begin{aligned}
\mathbb{E}[(e^2 - h(X))^2] &= \mathbb{E}[(e^2 - \sigma^2(X) + \sigma^2(X) - h(X))^2] \\
&= \mathbb{E}[(e^2 - \sigma^2(X))^2 + 2(e^2 - \sigma^2(X))(\sigma^2(X) - h(X)) + (\sigma^2(X) - h(X))^2] \\
&= \mathbb{E}[(e^2 - \sigma^2(X))^2] + 2\mathbb{E}[(e^2 - \sigma^2(X))(\sigma^2(X) - h(X))] + \mathbb{E}[(\sigma^2(X) - h(X))^2] \\
&= \mathbb{E}[(e^2 - \sigma^2(X))^2] + \mathbb{E}[(\sigma^2(X) - h(X))^2] \\
&\geq \mathbb{E}[(e^2 - \sigma^2(X))^2].
\end{aligned}$$

This holds for *any* predictor $h(X)$. Therefore the MSQE is minimized when $h(X) = \sigma^2(X)$.

The last equality follows from the L.I.E., together with the definition of $\sigma^2(X) = \mathbb{E}_X[e^2|X]$:

$$\begin{aligned}
\mathbb{E}[(e^2 - \sigma^2(X))(\sigma^2(X) - h(X))] + \mathbb{E}[(\sigma^2(X) - h(X))^2] &= \mathbb{E}[\mathbb{E}_X[(e^2 - \sigma^2(X))(\sigma^2(X) - h(X))|X]] \\
&= \mathbb{E}[(\mathbb{E}_X[e^2|X] - \sigma^2(X))(\sigma^2(X) - h(X))] \\
&= \mathbb{E}[(\sigma^2(X) - \sigma^2(X))(\sigma^2(X) - h(X))] \\
&= 0.
\end{aligned}$$

□

---

9. [2.21, Hansen] *Consider the short and long projections*

$$Y = X\gamma_1 + e,$$

$$Y = X\beta_1 + X^2\beta_2 + u.$$

(a) *Under what condition does $\gamma_1 = \beta_1$?*

Solution. From the analysis of omitted variable bias, we know that $\gamma_1 = \beta_1$ under one of two conditions: $\beta_2 = 0$ in the long regression or $\mathbb{E}[x_i x_i^2] = \mathbb{E}[x_i^3] = 0$. Note that if $\mathbb{E}[x_i] = 0$, the latter condition is equivalent to x_i having zero skewness. □

(b) *Take the long projection as $Y = X\theta_1 + X^3\theta_2 + \eta$. Is there a condition under which $\gamma_1 = \theta_1$?*

Solution. From the same argument, $\gamma_1 = \theta_1$ under one of two conditions: $\theta_2 = 0$ in the long regression, or $\mathbb{E}[x_i x_i^3] = \mathbb{E}[x_i^4] = 0$. The latter condition is impossible. Therefore $\gamma_1 = \theta_1$ only if $\theta_2 = 0$ in the long regression. □
