Daeyoung Lim
Grad School
Year of 2007, early

1. **(1)** $X_1, X_2, \ldots, X_n$ are a random sample whose p.m.f. is $\mathbb{P}(X = x) = (1-\theta)^x \theta,\ x = 0, 1, 2, \cdots$. What is the distribution of $\min(X_1, X_2, \ldots, X_N)$?
(2) $X_1, X_2, \ldots, X_n$ are independent random variables, of which the distribution of X_i is an exponential distribution whose mean is μ_i $(i = 1, 2, \ldots, n)$. What is the distribution of $\min(X_1, X_2, \ldots, X_n)$?

Solution: (1) We start with the c.d.f. of $\min(X_1, X_2, \ldots, X_n)$. For notational convenience, let us write $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$.

$$\begin{aligned}
\mathbb{P}\left(X_{(1)} \leq x\right) &= 1 - \mathbb{P}(X_1 > x, X_2 > x, \ldots, X_n > x) \\
&= 1 - \left(1 - \sum_{i=0}^{x} (1-\theta)^i \theta\right)^n \\
&= 1 - \left(1 - \theta \frac{1 - (1-\theta)^{x+1}}{1 - (1-\theta)}\right)^n \\
&= 1 - (1-\theta)^{n(x+1)}
\end{aligned}$$

Recall we are dealing with a discrete distribution. To get the p.m.f. from the c.d.f., we subtract the $(x-1)^{\text{th}}$ term from the x^{th} term.

$$\begin{aligned}
F(x) - F(x-1) &= 1 - (1-\theta)^{n(x+1)} - 1 + (1-\theta)^{nx} \\
&= (1-\theta)^{nx} - (1-\theta)^{n(x+1)} \\
&= (1-\theta)^{nx}\left(1 - (1-\theta)^n\right) \\
&\sim \text{Geo}\left(1 - (1-\theta)^n\right)
\end{aligned}$$

It turned out to be the geometric distribution whose random variable X indicates the number of failures before observing the first success.

Solution: (2) The p.d.f. of the i^{th} random variable is $f_{X_i}(x) = \frac{1}{\mu_i} e^{-x/\mu_i}$. Therefore,

$$\begin{aligned}
\mathbb{P}\left(X_{(1)} \leq x\right) &= 1 - \mathbb{P}(X_1 > x)\,\mathbb{P}(X_2 > x) \cdots \mathbb{P}(X_n > x) \\
&= 1 - \prod_{i=1}^{n} \int_x^{\infty} \frac{1}{\mu_i} e^{-x_i/\mu_i}\, dx_i \\
&= 1 - \prod_{i=1}^{n} e^{-x/\mu_i} \\
&= 1 - \exp\left(-\left(\frac{x}{\mu_1} + \frac{x}{\mu_2} + \cdots + \frac{x}{\mu_n}\right)\right) \\
&= 1 - \exp\left(-x \sum_{i=1}^{n} \frac{1}{\mu_i}\right)
\end{aligned}$$

Since the distribution is continuous, we can differentiate the c.d.f. to compute the p.d.f..

$$\begin{aligned}\frac{d}{dx}\mathbb{P}\left(X_{(1)} \le x\right) &= f_{X_{(1)}}\left(x\right) \\ &= \left(\sum_{i=1}^{n}\frac{1}{\mu_i}\right)\exp\left(-x\sum_{i=1}^{n}\frac{1}{\mu_i}\right) \\ &\sim \text{Exp}\left(\sum_{i=1}^{n}\frac{1}{\mu_i}\right)\end{aligned}$$

The minimum follows an exponential distribution whose mean is $\left(\sum_{i=1}^{n}\mu_i^{-1}\right)^{-1}$.

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathcal{N}\left(\mu, \sigma^2\right)$.
(1) If μ, σ^2 are unknown, explain the MLE method for μ, σ^2.
(2) Is the MLE for σ^2 the MVUE?

Solution: (1) The method of maximum likelihood attempts to find the optimal point in the domain that maximizes the likelihood function. According to optimization theory, we can find the optimal point of a function by differentiating it and equating it with 0 if the function is differentiable within the given domain. It is possible to take the derivative with respect to the parameter of interest directly to the function. However, for computational convenience, it is more commonplace to take the logarithm before differentiation on the ground that a logarithmic function is a monotically increasing function. Finally, to check whether the calculated optimum is a minimum or a maximum, we take the second derivative and see if it is positive or negative in the given region.

This methodology is also applicable to a normal distribution with unknown mean and variance.

$$\begin{aligned}L\left(\mu,\sigma^2\right) &= \left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\} \\ \ln L\left(\mu,\sigma^2\right) &= -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 \\ \frac{\partial}{\partial\mu}\ln L\left(\mu,\sigma^2\right) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right) = 0 \qquad \cdots(1) \\ \frac{\partial}{\partial\sigma^2}\ln L\left(\mu,\sigma^2\right) &= -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 = 0 \qquad \cdots(2)\end{aligned}$$

From eqn (1), we can immediately find out that $\hat{\mu}^{\text{MLE}} = n^{-1}\sum_{i=1}^{n}X_i$, which is the sample mean. Further, from eqn (2), it is obvious that $\widehat{\sigma^2}^{\text{MLE}} = n^{-1}\sum_{i=1}^{n}\left(X_i-\mu\right)^2$. Since this is a two-dimensional parameter space, we can plug in the MLE for μ to the second result thereby yielding $n^{-1}\sum_{i=1}^{n}\left(X_i-\hat{\mu}^{\text{MLE}}\right)^2$. (We will not go through the second derivatives.)

Solution: **(2)** *MVUE* stands for the *minimum variance unbiased estimator.* It is a term that describes an estimator in the class of all unbiased estimators that have the smallest variance. Recall the *bias-variance trade-off.* If we drop the 'unbiasedness' requirement , then apparently, the variance becomes smaller; there is no point in comparing an unbiased estimator and a biased one due to this nature. It is a well-known fact that an uncorrected sample variance is biased with an expected value of $n^{-1}(n-1)\sigma^2$ regardless of what distribution each random variable follows. Let us write $\overline{X}_n$ for $\hat{\mu}^{\text{MLE}}$.

$$
\begin{aligned}
\mathbb{E}\left[\widehat{\sigma}^2\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2-2X_i\overline{X}_n+\overline{X}_n^2\right] \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}\left(\mu^2+\sigma^2\right)-2n\mathbb{E}\left[\overline{X}_n^2\right]+n\mathbb{E}\left[\overline{X}_n^2\right]\right) \\
&= \frac{1}{n}\left(n\left(\mu^2+\sigma^2\right)-n\mathbb{E}\left[\overline{X}_n^2\right]\right) \\
&= \frac{1}{n}\left(n\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\frac{\sigma^2}{n}\right)\right) \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

Therefore, the MLE of σ^2 does not even count as a candidate to be compared its variance as it failed to become an unbiased estimator.

3. Independent random variables Y_i $(i=1,2,\ldots,n)$ each follows $\mathcal{N}\left(\beta_0+\beta_1 x_i,\sigma^2\right)$.

(1) Derive the maximum likelihood estimator(MLE) of the parameter σ^2.

(2) Show whether the estimator derived in **(1)** is biased or not.

Solution: **(1)** This is not very different from a linear regression setting. Let us first write down the likelihood function of Y_i.

$$
L\left(\beta_0,\beta_1,\sigma^2\right)=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\beta_0-\beta_1 x_i\right)^2\right\}
$$

As described in the previous answer, we will take the logarithm of the likelihood function and for simplicity, it will be denoted with $\ell\left(\beta_0, \beta_1, \sigma^2\right)$.

$$\begin{aligned}
\ell\left(\beta_0, \beta_1, \sigma^2\right) &= -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2 \\
\frac{\partial}{\partial\beta_0}\ell\left(\beta_0, \beta_1, \sigma^2\right) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right) = 0 \qquad \cdots (1) \\
\frac{\partial}{\partial\beta_1}\ell\left(\beta_0, \beta_1, \sigma^2\right) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)x_i = 0 \qquad \cdots (2) \\
\frac{\partial}{\partial\sigma^2}\ell\left(\beta_0, \beta_1, \sigma^2\right) &= -\frac{n}{2\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2 = 0 \qquad \cdots (3)
\end{aligned}$$

Rearranging eqn **(1)** with respect to β_0,

$$\hat{\beta}_0 = \overline{y}_n - \hat{\beta}_1\overline{x}_n.$$

Likewise with eqns **(2)**, **(3)**,

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} \\
\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2
\end{aligned}$$

Since $\hat{\beta}_0, \hat{\beta}_1$ depend on each other, we can use the two equations which produces

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} \\
&= \frac{\sum_{i=1}^{n} x_i y_i - \left(\overline{y}_n - \hat{\beta}_1\overline{x}_n\right)\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} \\
\left(1 - \frac{\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{\sum_{i=1}^{n} x_i^2}\right)\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - \overline{y}_n\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - \overline{y}_n\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}
\end{aligned}$$

We have just rediscovered that the least square estimator and the maximum likelihood estimator are identical under a linear regression. With these estimators for β_0, β_1, the MLE of σ^2 is $n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$.

Solution: (2) We all know the answer is a resounding *NO!* but since it is a problem, we will try to mathematically prove that it is a biased estimator. For convenience, we will introduce a matrix notation. A random variable with no subscript will from now on denote a random vector. For example, $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$. Then

$$\begin{aligned}
X &= \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \\
\hat{\beta} &= \begin{bmatrix} \hat{\beta}_0 & \hat{\beta}_1 \end{bmatrix}^T \\
\widehat{\sigma}^2 &= \frac{1}{n} \left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right)
\end{aligned}$$

The estimator $\hat{\beta}$ has the relation $\hat{\beta} = \left(X^T X\right)^{-1} X^T y$. Furthermore, $y \sim \mathcal{N}\left(X\beta, \sigma^2 I_n\right)$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}\right] &= \mathbb{E}\left[\left(X^T X\right)^{-1} X^T y\right] \\
&= \left(X^T X\right)^{-1} X^T \mathbb{E}\left[y\right] \\
&= \cancel{\left(X^T X\right)^{-1} X^T X} \beta \\
&= \beta
\end{aligned}$$

It is easy to show that the MLE for β is an unbiased one. That being said,

$$\mathbb{E}\left[\widehat{\sigma}^2\right] = \frac{1}{n}\mathbb{E}\left[\left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right)\right].$$

Let us write $H = X\left(X^T X\right)^{-1} X^T$ which is called the *hat matrix* in linear regression. We also call this the *projection matrix* and thus denote it by P in linear algebra. It allows us to write $X\hat{\beta} = Hy$.

$$\begin{aligned}
\mathbb{E}\left[\widehat{\sigma}^2\right] &= \frac{1}{n}\mathbb{E}\left[\left(y - Hy\right)^T \left(y - Hy\right)\right] \\
&= \frac{1}{n}\mathbb{E}\left[y^T \left(I_n - H\right)^T \left(I_n - H\right) y\right] \\
&= \frac{1}{n}\mathbb{E}\left[y^T \left(I_n - H\right) y\right] \quad \left(\because \left(I_n - H\right) = \left(I_n - H\right)^T = \left(I_n - H\right)^2\right) \\
&= \frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left(I_n - H\right) yy^T\right] \quad \left(\because y^T \left(I_n - H\right) y \text{ is a scalar.}\right) \\
&= \frac{1}{n}\,\mathrm{Tr}\left(\left(I_n - H\right)\mathbb{E}\left[yy^T\right]\right) \quad \left(\because \mathrm{Tr}, \mathbb{E} \text{ are linear.}\right) \\
&= \frac{1}{n}\,\mathrm{Tr}\left(\left(I_n - H\right)\left(\sigma^2 I_n + X\beta\beta^T X^T\right)\right) \\
&= \frac{1}{n}\left(\mathrm{Tr}\left(\sigma^2 I_n\right) + \mathrm{Tr}\left(X\beta\beta^T X^T\right) - \mathrm{Tr}\left(\sigma^2 H\right) - \mathrm{Tr}\left(HX\beta\beta^T X^T\right)\right)
\end{aligned}$$

$HX\beta\beta^T X^T = X\cancel{\left(X^T X\right)^{-1} X^T X}\beta\beta^T X^T = X\beta\beta^T X^T$. Therefore, the result is

$$\begin{aligned}
\mathbb{E}\left[\widehat{\sigma}^2\right] &= \frac{1}{n}\left(\sigma^2 \operatorname{Tr}\left(I_n\right) - \sigma^2 \operatorname{Tr}\left(H\right)\right) \\
&= \frac{1}{n}\left(n\sigma^2 - \sigma^2 \operatorname{Tr}\left(\left(X^T X\right)^{-1} X^T X\right)\right) \\
&= \frac{1}{n}\left(n\sigma^2 - \left(p+1\right)\sigma^2\right) \\
&= \frac{n-p-1}{n}\sigma^2
\end{aligned}$$

where p is the number of predictor variables. In our case $p = 1$. Thus, the expectation is $n^{-1}\left(n-2\right)\sigma^2$, which is biased.

4. **(1)** The following table defines the p.m.f. of a random variable X under a null hypothesis($\theta = \theta_0$) and an alternative hypothesis($\theta = \theta_1$). Compute the powers of the tests at $\alpha = 0.05$ significance level and derive the *most powerful test* via a comparison of the powers.

X	1	2	3	4	5	6
$f\left(x\vert\theta_0\right)$	0.01	0.02	0.02	0.05	0.10	0.80
$f\left(x\vert\theta_1\right)$	0.03	0.05	0.05	0.10	0.10	0.67

(2) Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathcal{N}\left(\mu, 1\right)$.

(a) If there exists a *most powerful test* for a null hypothesis $H_0 : \mu = 1$ and an alternative hypothesis $H_1 : \mu = 2$, derive it at 0.05 significance level and compute the power of the test used.

(b) If there exists a *most powerful test* for a null hypothesis $H_0 : \mu = 1$ and an alternative hypothesis $H_1 : \mu \neq 1$, derive it at 0.05 significance level. Explain if it does not exist.

(c) For the hypotheses given in the previous problem, compute the likelihood ratio test at 0.05 significance level and explain what optimal properties the method of testing has.

Solution: **(1)** Denote the rejection region (or equivalently the critical region) by RR. Then at 0.05 significance level, there are two rejection regions whose probabilities add up 0.05: $\mathrm{RR}_1 = \{1, 2, 3\}$, $\mathrm{RR}_2 = \{4\}$.

$$\Pr\left(\mathrm{RR}_1 \vert H_0\right) = \Pr\left(\mathrm{RR}_2 \vert H_0\right) = 0.05. \tag{1}$$

The power of each test is as follows:

$$\Pr\left(\mathrm{RR}_1 \vert H_1\right) = 0.13 \tag{2}$$

$$\Pr\left(\mathrm{RR}_2 \vert H_1\right) = 0.1. \tag{3}$$

Therefore, the most powerful test is rejecting H_0 when $\mathrm{RR} = \{1, 2, 3\}$.

Solution: (2)

- $H_0 : \mu = 1 \quad H_1 : \mu = 2$. By Neyman-Pearson lemma, the most powerful test is derived by the likelihood ratio $\Lambda = L_0/L_1$.

$$\begin{aligned}
\Lambda &= \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(x_i-1\right)^2\right)}{\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(x_i-2\right)^2\right)} &(4)\\
&= \exp\left(\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - 4\sum_{i=1}^{2} x_i + 4n\right) - \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i + n\right)\right) &(5)\\
&= \exp\left(-\sum_{i=1}^{n} x_i + \frac{3}{2}n\right) &(6)
\end{aligned}$$

  Therefore, reject H_0 when $\exp\left(-\sum_{i=1}^{n} x_i + \frac{3}{2}n\right) \leq k$.

  We see that the rejection region is when $\overline{X}_n \geq c$ for some constant c. Since $\sqrt{n}\left(\overline{X}_n - \mu\right) \sim \mathcal{N}(0,1)$, we can use the identity with the significance level to get the critical value.

$$\begin{aligned}
\alpha &= \Pr\left(\mathrm{RR}|H_0\right) &(7)\\
&= \Pr\left(\sqrt{n}\left(\overline{X}_n - 1\right) \geq z_{1-\alpha}\right) &(8)\\
&= \Pr\left(\overline{X}_n \geq \frac{z_{1-\alpha}}{\sqrt{n}} + 1\right) &(9)
\end{aligned}$$

  where $z_{1-\alpha}$ is the $(1-\alpha)^{\text{th}}$ quantile of standard normal. So $\overline{X}_n \geq n^{-1/2}z_{1-\alpha} + 1$ is the rejection region. To get the power,

$$\Pr\left(\mathrm{RR}|H_1\right) = \Pr\left(\sqrt{n}\left(\overline{X}_n - 2\right) \geq z_{1-\alpha} - \sqrt{n}\right). \tag{10}$$

- The likelihood ratio is as follows:

$$\frac{L_0}{L_1} = \exp\left(\left(1-\mu_1\right)\sum_{i=1}^{n} x_i + \frac{1}{2}\left(\mu_1^2 - n\right)\right) \leq k \tag{11}$$

  So when $\mu_1 > 1$, the rejection region becomes $\overline{X}_n \geq c_1$ and when $\mu_1 < 1$ then the rejection region is $\overline{X}_n \leq c_2$. Therefore, most powerfull tests exist for both cases but because the rejection regions depend on the parameter of the alternative hypothesis, uniformly powerful test does not exist.

- $\hat{\mu}^{\text{MLE}} = \overline{X}_n$. Therefore,

$$\begin{aligned}
\frac{\sup L_0}{\sup L} &= \exp\left(-\frac{n}{2}\left(\overline{X}_n^2 - 2\overline{X}_n + 1\right)\right) &(12)\\
&= \exp\left(-\frac{n}{2}\left(\overline{X}_n - 1\right)^2\right) \leq k. &(13)
\end{aligned}$$

  $\mathrm{RR} = \left\{\overline{X}_n \leq c_1 \text{ or } \overline{X}_n \geq c_2\right\}$ which becomes

$$\begin{aligned}
\alpha &= \Pr\left(\mathrm{RR}|H_0\right) &(14)\\
&= \Pr\left(\left|\sqrt{n}\left(\overline{X}_n - 1\right)\right| \geq z_{\alpha/2}\right) &(15)
\end{aligned}$$