# Statistics 5101, Fall 2000, Geyer
## Homework Solutions #11

## Problem L7-1

**(a)**

$$
\begin{aligned}
f_X(x) &= \prod_{i=1}^n f(x_i) \\
&= \prod_{i=1}^n \frac{1}{b-a} \\
&= \frac{1}{(b-a)^n}, \qquad\qquad a < x_i < b
\end{aligned}
$$

**(b)**

$$
\begin{aligned}
f_X(x) &= \prod_{i=1}^n f(x_i) \\
&= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\right\}
\end{aligned}
$$

**(c)**

$$
\begin{aligned}
f_X(x) &= \prod_{i=1}^n f(x_i) \\
&= \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} \\
&= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, \qquad\qquad x_i = 0, 1
\end{aligned}
$$

**(d)**

$$
\begin{aligned}
f_X(x) &= \prod_{i=1}^n f(x_i) \\
&= \prod_{i=1}^n \frac{m^{x_i} e^{-m}}{x_i!} \\
&= \frac{m^{\sum_{i=1}^n x_i} e^{-nm}}{\prod_{i=1}^n x_i!}, \qquad\qquad x_i = 0, 1, 2, ..
\end{aligned}
$$

**(e)**

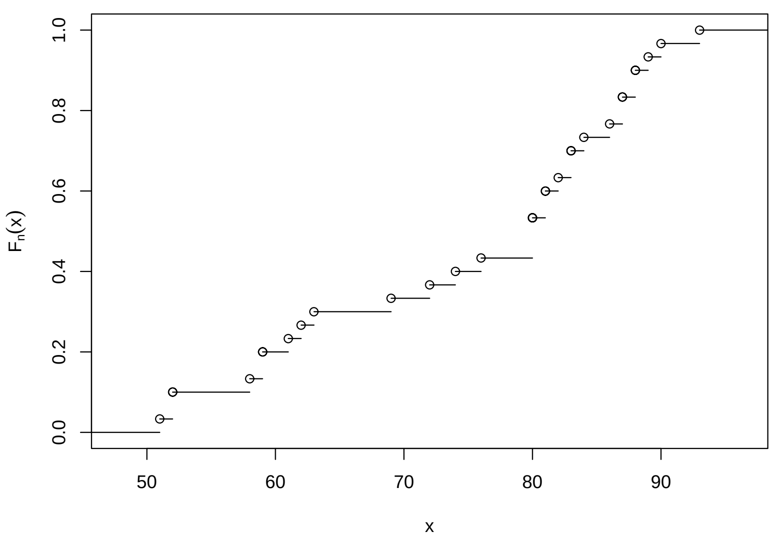
$$\begin{aligned} f_X(x) &= \prod_{i=1}^{n} f(x_i) \\ &= \prod_{i=1}^{n} (1-p)^{x_i} p \\ &= (1-p)^{\sum_{i=1}^{n} x_i} p^n, \quad x = 0, 1, 2, .. \end{aligned}$$

**(f)**

$$\begin{aligned} f_X(x) &= \prod_{i=1}^{n} f(x_i) \\ &= \prod_{i=1}^{n} \frac{1/\pi}{1+(x_i-\theta)^2} \\ &= \frac{1}{\pi^n} \frac{1}{\prod_{i=1}^{n} 1+(x_i-\theta)^2} \end{aligned}$$

# Problem L7-4

**(a)**



**(b)**

$$\begin{aligned}\overline{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ &= \frac{2250}{30} \\ &= 75\end{aligned}$$

$$\begin{aligned}S &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} \\ &= \sqrt{\frac{4888}{29}} \\ &= 12.98\end{aligned}$$

**(c)** The two middle values (in sorted order) are both 80. Hence the median is 80.

**(d)** There are $n = 30$ data points.

To find the $p$-th quantile of the empirical distribution, use Theorem 7.5 in the notes. For the lower quartile, $p = 1/4$. Then $np = 7.5$ is not an integer, and the lower quartile is data point number $\lceil np \rceil = 8$ (in sorted order), which is 62.

For the upper quartile, $p = 3/4$. Then $np = 22.5$ is not an integer, and the lower quartile is data point number $\lceil np \rceil = 23$ (in sorted order), which is 86.

**(e)** The interquartile range is $86 - 62 = 24$.

# Problem L7-6

**(a)**

$$\sum X = n \times \overline{X} = 10 \times 5 = 50$$

$$\sum X^2 = n\overline{X}^2 + (n-1)S_X^2 = 10 \times 5^2 + 9 \times 2^2 = 286.$$

**(b)**

$$\overline{Y} = 3 + 4\overline{X} = 3 + 4 \times 5 = 23$$

$$S_Y = 4S_X = 4 \times 2 = 8$$

## Problem L7-13

$$\overline{X} \sim N(\mu, .16)$$

$$P(|\overline{X} - \mu| > .8m) = P(|Z| > 2) = 2\Phi(-2) = .0456$$

## Problem L7-14

By the corollary on p. 210 in Lindgren we know that

$$E(S^2) = \sigma^2 = \frac{(1-0)^2}{12} = \frac{1}{12}.$$

In order to find the variance of $S^2$ we need to find $\mathrm{var}(V)$ using formula (5) p. 210 in Lindgren. Since

$$\mu_2 = \sigma_X^2 = \frac{1}{12},$$

and

$$\mu_4 = \int_0^1 (x - \frac{1}{2})^4 \, dx = \frac{1}{80}$$

we have

$$\mathrm{var}(V) = \frac{1}{180n} + \frac{1}{360n^2} - \frac{1}{120n^3}$$

and

$$\mathrm{var}(S^2) = \frac{n^2}{(n-1)^2}\,\mathrm{var}(V) = \frac{n^2}{(n-1)^2}\left[\frac{1}{180n} + \frac{1}{360n^2} - \frac{1}{120n^3}\right]$$

Sorry. The answer in the back of the book is wrong. It gives $\mathrm{var}(V)$ rather than $\mathrm{var}(S^2)$.

## Problem L7-16

**(a)**

$$E(D) = E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = 80 - 80 = 0$$

**(b)**

$$\mathrm{var}(D) = \mathrm{var}(\overline{X} - \overline{Y}) = \mathrm{var}(\overline{X}) + \mathrm{var}(\overline{Y}) = \frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y} = \frac{36}{100} + \frac{36}{150} = .6$$

**(c)** This cannot be done exactly, since the exact population distribution is not specified. By the central limit theorem $D \approx N(\mu_D, \sigma_D^2)$, that is, $D \approx N(0, .6)$. So we use that.

$$\begin{aligned} P(|D| > 2) &= 1 - P(-2 < D < 2) \\ &= 1 - P(-2.582 < Z < 2.582) \\ &= 2\Phi(-2.582) \\ &= 0.0098 \end{aligned}$$

# Problem L7-22

$$E(D) = 0$$

and

$$\operatorname{var}(D) = 36 \left( \frac{1}{10} + \frac{1}{15} \right) = 6$$

So $D \sim N(0, 6)$, and

$$\begin{aligned} P(|D| > 2) &= 1 - P(-2 < D < 2) \\ &= 1 - P(-0.8165 < Z < 0.8165) \\ &= 2\Phi(-0.8165) \\ &= .414 \end{aligned}$$

# Problem L7-25

$$E\left( \frac{\overline{X}_n - \mu}{S_n} \right) = E\left( \frac{1}{S_n} \right) E(\overline{X}_n - \mu)$$

by the independence of $S_n$ and $\overline{X}_n$ (Corollary to Theorem 10 of Chapter 7 in Lindgren or Theorem 7.24 in the notes) assuming the expectations exist. We know the second expectation exists (because $\overline{X}_n$ is normal and the normal distribution has first moments) and is equal to zero because the first central moment is always zero (Theorem 2.9 in the notes).

Thus the expectation is zero if $E(1/S_n)$ exists. To prove that we need to look at the density of the gamma distribution of $S_n^2$, Equation (7.34) in the notes. Write $Y_n = S_n^2$ so

$$Y_n \sim \operatorname{Gam}\left( \frac{n-1}{2}, \frac{n-1}{2\sigma^2} \right)$$

Then $S_n = \sqrt{Y_n}$ so the existence question is: Does $E(Y_n^{-1/2})$ exist when $Y_n$ has this gamma distribution? This existence problem was actually done as an example in the notes (Example 2.5.6). The result there is that if $X \sim \operatorname{Gam}(\alpha, 1)$

then $X^p$ has expectation whenever $p > -\alpha$. The same analysis applied here says that $E(1/S_n)$ exists if

$$p = -\frac{1}{2} > -\alpha = -\frac{n-1}{2}$$

that is, if $n > 2$. So $E\left\{(\overline{X}_n - \mu)/S_n\right\}$ is zero whenever $n > 2$ (and otherwise does not exist).

# Problem N6-1

We need to show that if

$$Y_n = X_1 + X_2 + \ldots + X_n \sim \text{Cauchy}(n\mu, n\sigma)$$

then

$$\frac{1}{n}Y_n = \overline{X}_n \sim \text{Cauchy}(n\mu, n\sigma)$$

Thus this is simply a question about the linear change of variables

$$\overline{X}_n = \frac{1}{n}Y_n$$

The linear change of variable theorem (Theorem 7 of Chapter 3 in Lindgren) says

$\bar{X}_n = Y_n/n$, by the transformation theorem, the Jacobin J is $n$, and the density of $\bar{X}_n$ is

$$\begin{aligned} f_{\overline{X}_n}(x) &= n f_{Y_n}(nx) \\ &= n\frac{n\sigma}{\pi(n^2\sigma^2 + n^2[x-\mu]^2)} \\ &= \frac{\sigma}{\pi(\sigma^2 + [x-\mu]^2)} \end{aligned}$$

which is what was to be shown.

# Problem N6-2

By the CLT

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} Y$$

where $Y \sim \mathcal{N}(0, \sigma^2)$. And by assumption

$$S_n \xrightarrow{P} \sigma.$$

Put these together using Slutsky's theorem. The function

$$g(u, v) = \frac{u}{v}$$

is continuous except at $v = 0$ hence is continuous a every point of the form $(u, \sigma)$ because $\sigma > 0$ by assumption.

Hence

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{\mathcal{D}} \frac{1}{\sigma} Y$$

and $Y/\sigma$ is standard normal by the rules for linear transformation of normal random variables.

# Problem N6-3

The $Y_n = I_A(X_n)$ are i. i. d. because functions of independent random variables are independent (Theorem 13 of Chapter 3 in Lindgren). The LLN says

$$\overline{Y}_n \xrightarrow{P} \mu_Y$$

so we only need to show that $\mu_Y = P(A)$, but this is obvious because "probability is just expectation of indicator functions" (Section 2.6 in the notes).

# Problem N6-4

To do this problem, we need to recognize that the $Y_n$ defined in the previous problem are $\text{Ber}(p)$ random variables. Every zero-one valued random variable $X$ is Bernoulli with "success" probability $p = P(X = 1)$. Every indicator function is zero-one valued, and $P(I_A = 1) = P(A)$ by definition ("probability is just expectation of indicator functions" again).

Therefore

$$\begin{aligned} E(Y_i) &= p \\ \text{var}(Y_i) &= p(1 - p) \end{aligned}$$

and the LLN and CLT say

$$\overline{Y}_n \xrightarrow{P} p$$

and

$$\sqrt{n}(\overline{Y}_n - p) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, p(1 - p)\big)$$

Now

$$g(u) = \sqrt{u(1 - u)}$$

is a continuous function, so the continuous mapping theorem applied to the LLN says

$$\sqrt{\overline{Y}_n(1 - \overline{Y}_n)} \xrightarrow{P} \sqrt{p(1 - p}$$

if we call the left hand side $S_n$ and the right hand side $\sigma$ (it is the standard deviation of the $Y_i$), then apply Problem 6.2 to this statement and the CLT, we get what was to be shown.