

Partial Solutions for h4/2014S: Sampling Distributions

- (1) Let X_1 and X_2 be two independent random variables, each with the same probability distribution given as follows.

$$f(x) = \frac{1}{2}e^{-x/2}, \quad x \geq 0$$

- (a) Compute the probability distribution function of the new random variable $Y = X_1 + X_2$.
- (b) What type of probability distribution is your answer in (a)?

S1(a) $\phi_1(t) = \phi_2(t) = \frac{1}{1-2t}$, for $t < \frac{1}{2}$.

$$\begin{aligned} E \left[e^{t(X_1+X_2)} \right] &= \int_0^\infty \int_0^\infty e^{t(x_1+x_2)} f(x_1)f(x_2)dx_1dx_2 \\ &= \left(\int_0^\infty e^{tx_1} f(x_1)dx_1 \right) \left(\int_0^\infty e^{tx_2} f(x_2)dx_2 \right) \\ &= \frac{1}{(1-2t)^2} \end{aligned}$$

Thus, $f_Y(y) = \frac{1}{4}ye^{-y/2}$, $0 < y < \infty$

- S1(b)** $Y = X_1 + X_2$ has a gamma distribution with parameters $\alpha = 2$ and $\theta = 2$ since $M_Y(t) = M_{X_1}(t) \times M_{X_2}(t) = \frac{1}{1-2t} \times \frac{1}{1-2t} = \frac{1}{(1-2t)^2}$.

- (2) Let the moment-generating function of X be

$$\phi(t) = e^{3t+2t^2}, \quad -\infty < t < \infty$$

- (a) Find the mean $E(X)$ and variance $\text{Var}(X)$.
- (b) Name the distribution of X.
- (c) Write down the p.d.f $f(x)$ for X.

S2(a) $E(X) = 3$ and $\text{Var}(X) = 4$.

S2(b) $X \sim N(3, 4)$, that is, X has a normal distribution with mean 3 and variance 4.

S2(c) The p.d.f of X is $f(x) = \frac{1}{\sqrt{8\pi}}e^{-(x-3)^2/8}$, $-\infty < x < \infty$.

- (3) Let $X_1 \sim b(n_1, p)$ and $X_2 \sim b(n_2, p)$ be independent r.v.s. Define $Y = X_1 + X_2$.

- (a) What is $M_Y(t)$?
 (b) How is Y distributed?

S3(a) $M_Y(t) = M_{X_1}(t)M_{X_2}(t) = (1 - p + pe^t)^{n_1}(1 - p + pe^t)^{n_2} = (1 - p + pe^t)^{n_1+n_2}$

S3(b) $Y \sim b(n_1 + n_2, p)$.

(4) If $X \sim N(\mu, \sigma^2)$, show that $Y = (aX + b) \sim N(a\mu + b, a^2\sigma^2)$, $a \neq 0$.

S4 Without loss of generality (W.L.O.G.), let $a > 0$, then

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) &&= P(X \leq \frac{y-b}{a}) \\ &= \int_{-\infty}^{(y-b)/a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} dx \end{aligned}$$

then

$$f(y) = F'(y) = \frac{1}{\sqrt{2\pi}(a\sigma)} e^{-[y-(a\mu+b)]^2/[2(a\sigma)^2]}, \quad -\infty < y < \infty$$

(5) The joint probability density function for two continuous random variables X and Y is given as follows.

$$f(x, y) = \begin{cases} \beta xy & 0 < x < 2 \text{ and } 0 < y < 4 \\ 0 & \text{elsewhere} \end{cases}$$

- (a) Determine the constant β to make it a joint p.d.f.
 (b) Determine the conditional probability distribution function $f(x|y)$.
 (c) Compute the covariance between X and Y .

S5(a) $\beta = \frac{1}{16}$, $f_X(y) = \int_0^4 \frac{1}{16} xy dx = \frac{x}{2}$, for $0 < x < 2$, $f_Y(y) = \int_0^2 \frac{1}{16} xy dx = \frac{y}{8}$, for $0 < y < 4$.

S5(b) $f(x|y) = \frac{f(x,y)}{f(y)} = \frac{x}{2}$, $f(x, y) = \frac{xy}{16} = \frac{x}{2} \times \frac{y}{8} = f_X(x)f_Y(y)$.

S5(c) $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$, where

$$\mu_x = \int_0^2 x \cdot \frac{x}{2} dx = \frac{4}{3}$$

and

$$\mu_y = \int_0^4 y \cdot \frac{y}{8} dy = \frac{8}{3},$$

Thus

$$\text{Cov}(X, Y) = E[XY] - \mu_X \mu_Y = \int_0^2 \int_0^4 \frac{1}{16} x^2 y^2 dy dx - \frac{32}{9} = \frac{8 \times 64}{144} - \frac{32}{9} = 0.$$

Note that X and Y are *uncorrelated*.

- (6) Show that the sum of n independent Poisson random variables with respective means $\lambda_1, \lambda_2, \dots, \lambda_n$ is Poisson with mean $\lambda = \sum_{i=1}^n \lambda_i$.

S6 Let $Y = \sum_{i=1}^n X_i$, then

$$\begin{aligned} E[e^{tY}] &= E[e^{t \sum_{i=1}^n X_i}] &= \sum_{x_1, x_2, \dots, x_n} [e^{\sum_{i=1}^n (tx_i)}] f(x_1) f(x_2) \cdots f(x_n) \\ &= \sum_{x_1, x_2, \dots, x_n} \prod_{i=1}^n [e^{tx_i} f(x_i)] &= \prod_{i=1}^n \left(\sum_{x_i} e^{tx_i} f(x_i) \right) \\ &= \prod_{i=1}^n e^{\lambda_i (e^t - 1)} &= e^{[\sum_{i=1}^n \lambda_i] (e^t - 1)} \end{aligned}$$

Thus, Y has a Poisson distribution with mean $\lambda = \sum_{i=1}^n \lambda_i$.

- (7) Let $Z_i \sim N(0, 1)$, for $1 \leq i \leq 7$ and define $W = \sum_{i=1}^7 Z_i^2$. Find $P(1.69 < W < 14.07)$.

S7 $Y \sim \chi^2(7)$, then $P(1.69 < W < 14.07) = P(W < 14.07) - P(W \leq 1.69) = 0.95 - 0.025 = 0.925$.

- (8) Assume there are 100 observations, denoted by x_i , $1 \leq i \leq 100$, and each is drawn from a population with a continuous distribution in $[0, 2]$.

- Give the formula for the sample mean \bar{X} and sample variance S^2 in this problem.
- Compute the sample mean and sample variance for \bar{X} .
- Try to give an approximate distribution for \bar{X} as best as you can and explain your reason.
- Compute the probability that the sample mean value is larger than 1.02. Note that you can express your solution with the cumulative standard normal distribution function

$$\Phi(r) = \int_{-\infty}^r \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

S8(a) $\bar{X} = \frac{1}{100} \sum_{i=1}^{100} X_i$, $S^2 = \frac{1}{99} \sum_{i=1}^{100} (X_i - \bar{X})^2$.

S8(b) $E(X_i) = \int_0^2 (x \cdot \frac{1}{2}) dx = 1$, $Var(X_i) = E(X_i^2) - [E(X_i)]^2 = \int_0^2 (x^2 \cdot \frac{1}{2}) dx - 1 = \frac{1}{3}$,

then

$$E(\bar{X}) = \frac{1}{100} \sum_{i=1}^{100} E(X_i) = 1$$

$$Var(\bar{X}) = \frac{1}{10000} \sum_{i=1}^{100} Var(X_i) = \frac{1}{300}$$

S8(c) By C.L.T.

$$\frac{\bar{X} - E(\bar{X})}{\sqrt{\frac{1}{100} Var(X_1)}} \rightarrow N(0, 1) \text{ in probability.}$$

Then

$$\begin{aligned} P(\bar{X} > 1.02) &= P(10\sqrt{3}(\bar{X} - 1) > 10\sqrt{3}(1.02 - 1)) = 1 - \Phi(0.02 * 10\sqrt{3}) \\ &= \Phi(-0.3464) \approx 0.3645 \end{aligned}$$

(9) Let $random()$ be a pseudo random number generator which can randomly generate a real in $[0, 1)$. Let $X \sim N(0, 1)$, and $Y \sim N(3, 16)$. Give algorithms to sample (simulate) the distributions of X and Y , respectively.

S9 Let $R_i \sim U(0, 1)$ and $R = \sum_{i=0}^n R_i$, by C.L.T., we have

$$\frac{R - E(R)}{\sqrt{n Var(R_1)}} \rightarrow N(0, 1) \text{ in probability}$$

or

$$\frac{R/n - E(R_1)}{\sqrt{Var(R_1)/n}} \rightarrow N(0, 1) \text{ in probability}$$

In application, if we choose $n = 12$, then

$$\left(\sum_{i=0}^{12} R_i\right) - 6 \approx X \sim N(0, 1)$$

(10) Let X_1, X_2, \dots, X_{30} be a random sample of size 30 from a Poisson distribution with a mean $2/3$. Approximate

(a) $P(15 < \sum_{i=1}^{30} X_i \leq 22)$.

(b) $P(21 \leq \sum_{i=1}^{30} X_i < 27)$.

S10(a)

$$\begin{aligned}
P(15 < \sum_{i=1}^{30} X_i \leq 22) &= P(15.5 \leq \sum_{i=1}^{30} X_i \leq 22.5) \\
&= P\left(\frac{(15.5-20)}{\sqrt{30 \times (2/3)}} \leq \frac{\sum_{i=1}^{30} X_i - 20}{\sqrt{30 \times (2/3)}} \leq \frac{(22.5-20)}{\sqrt{30 \times (2/3)}}\right) \\
&= P\left(\frac{-4.5}{\sqrt{20}} \leq \frac{\sum_{i=1}^{30} X_i - 20}{\sqrt{20}} \leq \frac{2.5}{\sqrt{20}}\right) \\
&= \Phi\left(\frac{2.5}{\sqrt{20}}\right) - \Phi\left(\frac{-4.5}{\sqrt{20}}\right) = \Phi(0.5590) - \Phi(-1.0062) \\
&\approx 0.7088 - 0.1630 = \mathbf{0.5458}
\end{aligned}$$

S10(b)

$$\begin{aligned}
P(21 \leq \sum_{i=1}^{30} X_i < 27) &= P\left(\frac{(20.5-20)}{\sqrt{30 \times (2/3)}} \leq \frac{\sum_{i=1}^{30} X_i - 20}{\sqrt{30 \times (2/3)}} \leq \frac{(26.5-20)}{\sqrt{30 \times (2/3)}}\right) \\
&= P\left(\frac{0.5}{\sqrt{20}} \leq \frac{\sum_{i=1}^{30} X_i - 20}{\sqrt{20}} \leq \frac{6.5}{\sqrt{20}}\right) \\
&= \Phi\left(\frac{6.5}{\sqrt{20}}\right) - \Phi\left(\frac{0.5}{\sqrt{20}}\right) = \Phi(1.4534) - \Phi(0.1118) \\
&\approx 0.9259 - 0.5435 = \mathbf{0.3824}
\end{aligned}$$

(11) Let X_1, X_2, \dots, X_n be a random sample of an exponential distribution with mean θ , that is, $f(x) = \frac{1}{\theta}e^{-x/\theta}$, $x > 0$.

- (a) Show that $P(\min(X_1, X_2, \dots, X_n) > 2) = e^{-2n/\theta}$.
- (b) Show that $P(\max(X_1, X_2, \dots, X_n) > 2) = 1 - (1 - e^{-2/\theta})^n$.
- (c) Find a minimum n such that $P(\min(X_1, X_2, \dots, X_n) > 2) \leq 3\%$ when $\theta = 2$.
- (d) Find a minimum n such that $P(\max(X_1, X_2, \dots, X_n) > 2) \geq 90\%$ when $\theta = 2$.

S11(a)

$$\begin{aligned}
P[\min(X_1, X_2, \dots, X_n) > 2] &= \prod_{i=1}^n P(X_i > 2) = \prod_{i=1}^n \left[\int_2^{\infty} \frac{1}{\theta} e^{-x/\theta} dx \right] \\
&= \prod_{i=1}^n \left[e^{-2/\theta} \right] = e^{-\frac{2n}{\theta}}
\end{aligned}$$

S11(b)

$$\begin{aligned} P[\max(X_1, X_2, \dots, X_n) > 2] &= P(\text{at least one } X_i > 2) = 1 - \prod_{i=1}^n P(X_i \leq 2) \\ &= 1 - \prod_{i=1}^n \left(\int_0^2 \frac{1}{\theta} e^{-x/\theta} dx \right) = 1 - [1 - e^{-2/\theta}]^n \end{aligned}$$

S11(c) Find a minimum n such that $P(\min(X_1, X_2, \dots, X_n) > 2) = e^{-2n/\theta} = e^{-2n/2} \leq 0.03$, that is, $e^{-n} \leq 0.03$, or $n \geq \lceil -\ln(0.03) = 3.5066 \rceil = 4$

S11(d) Find a minimum n such that $P(\max(X_1, X_2, \dots, X_n) > 2) = 1 - (1 - e^{-2/\theta})^n = 1 - [1 - e^{-2/2}]^n \geq 0.9$, that is, $1 - [1 - e^{-1}]^n \geq 0.9$ or $[1 - e^{-1}]^n \leq 0.1$. In other words, we want to find n such that $(e^{-1} + (0.1)^{1/n}) \geq 1$, or $(0.1)^{1/n} \geq (1 - e^{-1})$, or $n \ln(0.1) \geq \ln(\frac{e-1}{e})$. Thus, $n \geq 6$.

(12) Let $W_1 < W_2 < \dots < W_n$ be the order statistics of a random sample of size n from the uniform distribution $U(0,1)$.

- (a) Find the probability density function of W_1 .
- (b) Find the probability density function of W_n .
- (c) Use the result of **(a)** to find $E(W_1)$.
- (d) Use the result of **(b)** to find $E(W_n)$.

Solution:

$$\begin{aligned} G_r(y) &= P(W_r \leq y) = \sum_{k=r}^n \binom{n}{k} [F(y)]^k [1 - F(y)]^{n-k} \\ &= \sum_{k=r}^{n-1} \binom{n}{k} [F(y)]^k [1 - F(y)]^{n-k} + [F(y)]^n \end{aligned}$$

Thus the p.d.f. of W_r could be derived as

$$g_r(y) = G_r'(y) = \frac{n!}{(r-1)!(n-r)!} [F(y)]^{r-1} [1 - F(y)]^{n-r} f(y), \quad a < y < b \quad (1)$$

In particular,

$$g_1(y) = n[1 - F(y)]^{n-1} f(y), \quad a < y < b$$

$$g_n(y) = n[F(y)]^{n-1} f(y), \quad a < y < b$$

- (1) If X_i has a $U(0,1)$ distribution, then $E(W_r) = \frac{r}{n+1}$. In particular, $g_1(y) = n(1-y)^{n-1}$, $0 < y < 1$, $E(W_1) = \frac{1}{n+1}$; $g_n(y) = ny^{n-1}$, $0 < y < 1$, $E(W_n) = \frac{n}{n+1}$.
- (2) If X_j has an exponential distribution with mean 2, then $g_1(y) = ne^{-ny}$, $y > 0$ and $E(W_1) = \frac{1}{n}$.

***(13)** One of the most popular distributions used to model the lifetimes of electric components is the Weibull distribution, whose probability density function is given by

$$f(t) = \alpha t^{\alpha-1} e^{-t^\alpha}, \quad \text{for } t > 0, \quad \alpha > 0.$$

Determine for which values of α the hazard function of a Weibull random variable is increasing, for which values it is decreasing, and for which values it is constant.

***S13** $f'(t) = \alpha(\alpha - 1)t^{\alpha-2}e^{-t^\alpha} - \alpha^2 t^{2\alpha-2}e^{-t^\alpha}$, thus,

$$\begin{aligned} f'(t) &> 0 \quad \text{if } t < \left(\frac{\alpha-1}{\alpha}\right)^{1/\alpha} \\ &< 0 \quad \text{if } t > \left(\frac{\alpha-1}{\alpha}\right)^{1/\alpha} \end{aligned}$$