NANYANG TECHNOLOGICAL UNIVERSITY

SPMS/DIVISION OF MATHEMATICAL SCIENCES

2016/17 Semester 1 MH2500 Probability and Introduction to Statistics Tutorial 5

For the tutorial on 15 September, let us discuss

- Ex. 2.5.52, 57, 60, 66, 70
- Ex. 3.8.2

Ex. 2.5.52. Suppose that in a certain population, individuals' heights are approximately normally distributed with parameters $\mu = 70$ and $\sigma = 3$ in.
a. What proportion of the population is over 6 ft. tall?
b. What is the distribution of heights if they are expressed in centimeters? In meters?
(Conversions: 1 inch = 2.54cm and 1 ft = 12 inches.)

[Solution:]
a. Let X $N(70, 3^2)$ and let Z $N(0, 1)$ be the standard normal. Then

$$P(X > 72) = P\left(Z > \frac{72-70}{3}\right) = 1 - P(Z \leq 0.667) = 1 - 0.7486 = 0.251.$$

Therefore, 25.1% of the population is over 6 feet.
b. Distribution in centimeters: $N(70 \times 2.54, (3 \times 2.54)^2) = N(177.8, 58.1)$.
Distribution in meters: $N(1.778, 0.00581)$.

Ex. 2.5.57. If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ where $a < 0$, show that $Y \sim N(a\mu + b, a^2\sigma^2)$.

[Solution:]
Method 1. For any real number y,

$$\begin{aligned} P(Y \leq y) &= P(aX + b \leq y) \\ &= P\left(X \geq \frac{y-b}{a}\right) \qquad \text{(Note the sign changed because } a < 0.) \\ &= 1 - F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\left(1 - F_X\left(\frac{y-b}{a}\right)\right) \\ &= -f_X\left(\frac{y-b}{a}\right)\frac{d}{dy}\left(\frac{y-b}{a}\right) \\ &= -\frac{1}{a}f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{(-a\sigma)\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y-b-a\mu}{-a\sigma}\right)^2\right]. \end{aligned}$$

Hence $Y \sim N(a\mu + b, (-a\sigma)^2) = N(a\mu + b, a^2\sigma^2)$.

Method 2. Let $Y = g(X) = aX + b$. Then $g^{-1}(y) = \frac{y-b}{a}$. By Proposition B (Handout 2 slide 57),

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= f_X\left(\frac{y-b}{a}\right) \left|\frac{1}{a}\right| \\ &= \frac{1}{(-a\sigma)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-b-a\mu}{-a\sigma}\right)^2\right] \qquad \text{(because } |1/a| = -1/a\text{)}. \end{aligned}$$

Ex. 2.5.60. Find the density function of $Y = e^Z$, where $Z \sim N(\mu, \sigma^2)$. This is called the **lognormal density**, since $\log Y$ is normally distributed.

[Solution:] Method 2. Let $Y = g(Z) = e^Z$. Then $y > 0$ for all values of z. Also, the inverse of g exists and is given by $g^{-1}(y) = \log y$. By Proposition B (Handout 2 slide 57),

$$\begin{aligned} f_Y(y) &= f_Z(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= f_Z(\log y) \left|\frac{1}{y}\right| \\ &= \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2\right]. \end{aligned}$$

Ex. 2.5.66. Let $f(x) = \alpha x^{-\alpha-1}$ for $x \geq 1$ and $f(x) = 0$ otherwise, where α is a positive parameter. Show how to generate random variables with this density from a uniform random number generator.

[Solution:] The cdf of F is

$$F(t) = \int_1^t \alpha x^{-\alpha-1} dx = [-x^{-\alpha}]_1^t = 1 - \frac{1}{t^\alpha}.$$

Therefore, we may find F^{-1} by solving $y = 1 - \frac{1}{t^\alpha}$ for t. This equation is equivalent to

$$\frac{1}{t^\alpha} = 1 - y \qquad \text{which implies that} \qquad t = \frac{1}{(1-y)^{1/\alpha}}.$$

Therefore, $F^{-1}(y) = 1/(1-y)^{1/\alpha}$.

Suppose U is uniform on $[0,1]$. Let $T = F^{-1}(U) = 1/(1-U)^{1/\alpha}$. Then by Proposition D (Handout 2 slide 59), T has cdf $1 - \frac{1}{t^\alpha}$, that is T is a random variable with density $f(t)$.

Ex. 2.5.70. Let U be a uniform random variable on $[0,1]$. Find the density function of $V = U^{-\alpha}$, $\alpha > 0$. Compare the rates of decrease of the tails of the densities as a function of α. Does the comparison make sense intuitively?

[Solution:]
Let U be a uniform random variable on $[0,1]$.

$$\begin{aligned} P(V \le v) &= P(U^{-\alpha} \le v) \\ &= P\left(U^\alpha \ge \frac{1}{v}\right) \qquad \text{(since } U^\alpha \ge 0\text{)} \\ &= P\left(U \ge \frac{1}{v^{1/\alpha}}\right) \\ &= 1 - \frac{1}{v^{1/\alpha}}. \end{aligned}$$

Therefore, the density function is

$$f_V(v) = \frac{d}{dv}\left(1 - \frac{1}{v^{1/\alpha}}\right) = \frac{1}{v^{1/\alpha+1}\alpha}, \qquad (1 \le v < \infty).$$

For a large α value, $1/\alpha$ is small and close to zero and so $1/v^{1/\alpha}$ tends to 1 from below. Thus the tails of the densities tends towards $1/(v\alpha)$ from below, and the larger the value of α, the faster the decrease.

Intuitively this makes sense. The larger the value of α, the more steep the graph of $g(U) = U^{-\alpha}$ will be for values of U near zero. The density of V involves $g^{-1}(V)$, and so the steeper graph of $g(U)$ near zero translates to a faster decrease in the tail of the density.

Ex. 3.8.2. An urn contains p black balls, q white balls, and r red balls; and n balls are chosen without replacement.

a. Find the joint distribution of the numbers of black , white, and red balls in the sample.
b. Find the joint distribution of the numbers of black and white balls in the sample.
c. Find the marginal distribution of the number of white balls in the sample.

[Solution:] Let X, Y, and Z denote the number of black, white, and red balls chosen.

a.

$$p(x,y,z) = \begin{cases} \frac{\binom{p}{x}\binom{q}{y}\binom{r}{z}}{\binom{p+q+r}{n}}, & \text{if } 0 \le x \le p,\ 0 \le y \le q,\ 0 \le z \le r \text{ and } x+y+z=n; \\ 0, & \text{otherwise.} \end{cases}$$

b.

$$p(x,y) = \begin{cases} \frac{\binom{p}{x}\binom{q}{y}\binom{r}{n-x-y}}{\binom{p+q+r}{n}}, & \text{if } 0 \le x \le p,\ 0 \le y \le q, \text{ and } 0 \le x+y \le n; \\ 0, & \text{otherwise.} \end{cases}$$

c.

$$p(y) = \begin{cases} \frac{\binom{q}{y}\binom{p+r}{n-y}}{\binom{p+q+r}{n}}, & \text{if } 0 \le y \le q \text{ and } 0 \le y \le n; \\ 0, & \text{otherwise.} \end{cases}$$