Ömer Oktay Gültekin
Section : 5
Department : CS

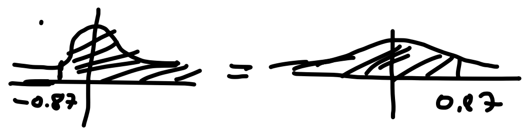
Note : I uploaded my homework 1 day late with the permission of Dilek Hoca, F41. ☺

1. Twelve percent of the population is left handed. Approximate probability that there at least 20 lefthanders in a school of 200 students.

x : # of lefthanders in a school of 200 students.
$X \sim Bin(n=200, p=0.12)$. To approximate, the normal app. to the Binomial dist. should be used $\Rightarrow np = 24$,
$\sqrt{npq} = \sqrt{24 \cdot \frac{88}{100}} = \frac{\sqrt{2112}}{10} \approx 4.596$ where np and $\sqrt{npq}$ is μ and σ, respectively.

$P(X \geq 20)$ is (without continuity correction) $\approx$

$$P\left(\frac{X-24}{4.596} \geq \frac{20-24}{4.596}\right) \approx P\left(Z \geq -0.87\right) =$$



$$\Rightarrow P(Z \geq -0.87) = P(Z \leq 0.87) = F(0.87)$$
$$= 0.8078 \Rightarrow \boxed{P(X \geq 20) \approx 0.8078}$$

2. Let V denote rainfall volume and W denote runoff volume (both in mm). According to the article "Runoff (*Journal of Water Resource Planning and Management*, 2006: 4-14), the runoff volume will be 0) if $V \leq v_d$ and will be k ($V - v_d$) if $V > v_d$. Here v_d is volume of depression storage (a constant) and k (also a constant. Cited article proposes exponential distribution with parameter λ for rainfall volume V. Find pdf of W.

$V \sim exp(\lambda) \Rightarrow f(v) = \lambda e^{-\lambda v}$ * $v > 0$

$$W = \begin{cases} 0 & V \leq V_d \\ k(V - V_d) & V > V_d \end{cases}$$

$P(V \leq V_d) = F_V(V_d) = 1 - e^{-\lambda V_d}$ where $V_d > 0$.
$\underline{V > V_d} \Rightarrow V - V_d > 0 \Rightarrow k(V - V_d) = \boxed{W > 0}$ //
$F_W(w) = P(W < w) = P(k(V - V_d) < w) = P(V < \frac{w}{k} + V_d)$
$= F_V(\frac{w}{k} + V_d)$

$$F'_w(w) = \frac{dF_w(w)}{dw} = \frac{d}{dw} F_v\left(\frac{w}{k} + \omega d\right) = \frac{1}{k} f_v\left(\frac{w}{k} + \omega d\right)$$

$$= \boxed{\frac{\lambda}{k} e^{-\lambda\left(\frac{w}{k} + \omega d\right)} = f_w(w)} \text{ using } *$$

$\underline{v \leq vd} \Rightarrow F_w(0) = P(w < 0) = P(0 < 0) = F_v(0)$, $F_w(0)' = \frac{d}{dw} F_v(0)$

$= 0 \Rightarrow$ pdf of w $\quad f(w) = \begin{cases} 0 & w \leq 0 \\ \frac{\lambda}{k} e^{-\lambda\left(\frac{w}{k} + \omega d\right)} & w > 0 \end{cases}$

3. The article " On Assessing the Accuracy of Offshore Wind Turbine Reliability - Based Design Loads from the Environmental Counter Method" (*Int. Journal of Offshore and Polar Engr.*, 2005: 132-140) proposes the Weibull distribution with α = 1.307 and β= 1.817 as a model for 1-hour significant wave height (m) at a certain site.

a) What is the probability that wave height is at most 0.5 m?

b) What is the median of the wave height distribution?

X: the wave height (m) at a certain site for 1-hour

$X \sim \text{Weibull}(\alpha = 1.307, \beta = 1.817)$

$f(x) = \alpha \beta x^{\beta - 1} e^{-\alpha x^\beta} \quad x > 0$

a) $P(X \leq 0.5) = ?$

$$P(X \leq 0.5) = \int_0^{0.5} \alpha \beta x^{\beta-1} e^{-\alpha x^\beta} dx = \left(-e^{-\alpha x^\beta}\right)\Big|_0^{0.5}$$

$$= -e^{-(1.307)(0.5)^{1.817}} + 1 \approx \boxed{0.31}$$

b) $\eta_{0.5} = ?$ $\quad 0.5 = \int_0^{\eta_{0.5}} \alpha \beta x^{\beta-1} e^{-\alpha x^\beta} dx = \left(-e^{-\alpha x^\beta}\right)\Big|_0^{\eta_{0.5}}$

$$= e^{-\alpha \cdot 0^\beta} - e^{-\alpha (\eta_{0.5})^\beta} = 0.5 \Rightarrow 1 - e^{-\alpha(\eta_{0.5})^\beta} = 0.5$$

$$\Rightarrow 0.5 = e^{-\alpha(\eta_{0.5})^\beta} \Rightarrow \ln(0.5) = -\alpha(\eta_{0.5})^\beta$$

$$\Rightarrow \eta_{0.5} = \left(\frac{\ln(0.5)}{-\alpha}\right)^{\frac{1}{\beta}} \Rightarrow \boxed{\eta_{0.5} \approx 0.705}$$

4. The time (in hours) required to repair a machine is a gamma random variable with parameters $\alpha = 2$ and $\beta = 2$.

a) Find the probability that the repair time of a machine exceeds 2 hours.

b) If there are independent 40 machines to be repaired in a shop, approximate the probability that for the majority of the machines, repair times exceed 2 hours.

X : the repair time (in hours) of a machine

$X \sim \text{gamma}(\alpha = 2, \beta = 2)$

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} = \frac{1}{4\Gamma(2)} x^{1} e^{-\frac{x}{2}} \quad x > 0$$

a) $$P(X>2) = 1 - P(X \le 2) = 1 - \int_0^2 \frac{1}{4\Gamma(2)} x . e^{-\frac{x}{2}} dx$$

$$\Gamma(2) = (2-1)! = 1 \Rightarrow 1 - \int_0^2 \frac{x . e^{-\frac{x}{2}}}{4} dx = 1 - \frac{1}{4}\int_0^2 x . e^{-\frac{x}{2}} dx \quad (*)$$

Recall that,

$$\int f(x) g'(x) dx = f(x) g(x) - \int f'(x) g(x) dx$$

$\Rightarrow$ Let $f(x) = x \Rightarrow f'(x) = 1$, $g'(x) = e^{-\frac{x}{2}} \Rightarrow g(x) = -2e^{-\frac{x}{2}}$

$\Rightarrow (*)$ becomes $\Rightarrow 1 - \frac{1}{4}\left[x.(-2e^{-\frac{x}{2}}) - \int 1.(-2e^{-\frac{x}{2}})dx\right]\Big|_0^2$

$$= 1 + \frac{1}{2}\left[x . e^{-\frac{x}{2}} - \int e^{-\frac{x}{2}} dx\right]\Big|_0^2 = 1 + \frac{1}{2}\left[x . e^{-\frac{x}{2}} - (-2e^{-\frac{x}{2}})\right]\Big|_0^2$$

$$= 1 + \frac{1}{2}\left[x . e^{-\frac{x}{2}} + 2e^{-\frac{x}{2}}\right]\Big|_0^2 = 1 + \frac{1}{2}\left[(x+2)e^{-\frac{x}{2}}\right]\Big|_0^2 = 1 + \frac{1}{2}\left[4e^{-1} - 2\right]$$

$$= 1 + 2e^{-1} - 1 = \boxed{\frac{2}{e}}$$

b) Y : # of machines that repair time exceed 2 hours in ind. 40 machines

$Y \sim Bin(n=40, p = \frac{2}{e})$ found in part A

$P(Y \ge 21) = ?$ $\quad \mu = np = \frac{80}{e}$, $\sigma = \sqrt{npq} = \sqrt{40 . \frac{2}{e} . (1 - \frac{2}{e})}$

Using the normal approximation to the Bin. Dist:

$$P(Y \ge 21) \sim P\left(\frac{Y - \mu}{\sigma} \ge \frac{21 - \frac{80}{e}}{\sqrt{40 . \frac{2}{e} . (1-\frac{2}{e})}}\right) \sim P(Z \ge -3.02)$$

$$= P(Z \le 3.02) \sim \boxed{0.9987}$$

5. The joint density of X and Y is

$$f(x,y)=\begin{cases} cye^{-x} & x>0, \quad 0<y<2 \\ 0 & otherwise \end{cases}$$

a) Find the value of constant c.

b) Find marginal density functions of X and Y. Are X and Y independent?

a) $\int_0^\infty \int_0^2 c e^{-x} y \, dy \, dx = c \int_0^\infty e^{-x} \left(\frac{y^2}{2}\right)\Big|_0^2 dx = c \int_0^\infty 2e^{-x} dx$

$= 2c \int_0^\infty e^{-x} dx = 2c \left(-e^{-x}\right)\Big|_0^\infty = 2c\,(0+e^0) = 2c = 1$

$\Rightarrow \boxed{c = \frac{1}{2}}$

b) $f_x(x) = \int_0^2 \frac{1}{2} y e^{-x} dy = \frac{e^{-x}}{2} \int_0^2 y\,dy = \frac{e^{-x}}{2}\left(\frac{y^2}{2}\right)\Big|_0^2$

$= \frac{2e^{-x}}{2} = \boxed{e^{-x} \quad 0<x<\infty}$. $f_x(x)$ is a pdf because

$\int_0^\infty e^{-x} = \left(-e^{-x}\right)\Big|_0^\infty = 1.$

$f_y(y) = \int_0^\infty \frac{1}{2} y e^{-x} dx = \frac{y}{2} \int_0^\infty e^{-x} dx = \frac{y}{2}\left(-e^{-x}\right)\Big|_0^\infty = \boxed{\frac{y}{2} \quad 0<y<2}.$

$f_y(y)$ is a pdf because $\int_0^2 \frac{y}{2} = \left(\frac{y^2}{4}\right)\Big|_0^2 = 1.$

x and y are independent $\Leftrightarrow f(x,y) = f_x(x).f_y(y).$

$f(x,y) = \frac{y e^{-x}}{2} = \frac{y}{2}.e^{-x} = f_y(y).f_x(x) \Rightarrow \boxed{x \text{ and } y \text{ independent.}}$

6. Joint density of (X,Y) is given by

$$f(x,y)=\begin{cases} x+y & 0<x<1 \quad 0<y<1 \\ 0 & elsewhere \end{cases}$$

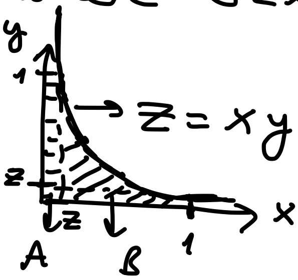
a) Find the density of $Z = XY$.

b) Find $Cov\,(X,Y)$.

a) $f_Z(z) = ?$

$$Z = xy \Rightarrow F_Z(z) = P(Z<z) = P(Xy<z) = \iint_{xy<z} f(x,y)\,dx\,dy$$

where $0<x<1$, $0<y<1$



$$\left.\begin{aligned} A &= \int_0^z \int_0^1 f(x,y)\,dy\,dx \\ B &= \int_z^1 \int_0^{z/x} f(x,y)\,dy\,dx \end{aligned}\right\} \Rightarrow A = \int_0^z \int_0^1 x+y \; dy\,dx$$

$$= \int_0^z \left(xy + \frac{y^2}{2}\right)\Big|_0^1 dx = \int_0^z x + \frac{1}{2}\,dx$$

$$= \left(\frac{x^2}{2} + \frac{x}{2}\right)\Big|_0^z = \frac{z^2}{2} + \frac{z}{2} = \boxed{\frac{z^2+z}{2}}$$

$$B = \int_z^1 \int_0^{z/x} (x+y)\,dy\,dx = \int_z^1 \left(xy + \frac{y^2}{2}\right)\Big|_0^{z/x} dx = \int_z^1 z + \frac{z^2}{2x^2}\,dx$$

$$= \left(xz - \frac{z^2}{2x}\right)\Big|_z^1 = \left(z - \frac{z^2}{2}\right) - \left(z^2 - \frac{z^2}{2z}\right) = \boxed{\frac{3z-3z^2}{2}}$$

Since $\max(xy) \leq 1$, $\min(xy) > 0$

$$A + B = \frac{z^2+z}{2} + \frac{3z-3z^2}{2} = \boxed{2z - z^2 = F_Z(z) \qquad 0<z<1}$$

$$\Rightarrow f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d(2z-z^2)}{dz} = 2-2z = f_Z(z),\ 0<z<1$$

$$\Rightarrow \text{pdf of } z,\ \boxed{f_Z(z) = \begin{cases} 2-2z & 0<z<1 \\ 0 & \text{elsewhere} \end{cases}}$$

b) $Cov(X,Y) = E(xy) - E(x)E(y)$, $E(g(x,y)) = \iint g(x,y) f(x,y)\,dy\,dx$

$$\Rightarrow E(x) = \int_0^1\int_0^1 x f(x,y)\,dy\,dx = \int_0^1\int_0^1 (x^2+xy)\,dy\,dx = \int_0^1 \left(x^2 y + \frac{xy^2}{2}\right)\Big|_0^1 dx$$

$$= \int_0^1 \left(x^2 + \frac{x}{2}\right)dx = \left(\frac{x^3}{3} + \frac{x^2}{4}\right)\Big|_0^1 = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}\,//$$

$$E(y) = \int_0^1 \int_0^1 y f(x,y)\, dy\, dx = \int_0^1 \int_0^1 (yx + y^2)\, dy\, dx$$

$$= \int_0^1 \left(\frac{y^2 x}{2} + \frac{y^3}{3} \right) \Big|_0^1 dx = \int_0^1 \left(\frac{x}{2} + \frac{1}{3} \right) dx = \left(\frac{x^2}{4} + \frac{x}{3} \right) \Big|_0^1 = \frac{1}{4} + \frac{1}{3} = \underline{\frac{7}{12}}$$

$$E(xy) = \int_0^1 \int_0^1 xy(x+y)\, dy\, dx = \int_0^1 \int_0^1 (x^2 y + y^2 x)\, dy\, dx$$

$$= \int_0^1 \left(\frac{x^2 y^2}{2} + \frac{y^3 x}{3} \right) \Big|_0^1 dx = \int_0^1 \left(\frac{x^2}{2} + \frac{x}{3} \right) dx = \left(\frac{x^3}{6} + \frac{x^2}{6} \right) \Big|_0^1$$

$$= \frac{1}{6} + \frac{1}{6} = \frac{4}{12} \Rightarrow \operatorname{COV}(x,y) = \frac{4}{12} - \frac{7}{12} \cdot \frac{7}{12} = \boxed{\frac{-1}{144} \approx -0.0069}$$

7. Let $\bar{x}_n$ and s_n^2 denote the sample mean and variance for the sample $x_1, x_2, \dots x_n$ and let $\bar{x}_{n+1}$ and s_{n+1}^2 denote those quantities when an additional observation x_{n+1} is added to the sample.

a) Show that $\bar{x}_{n+1}$ can be computed from $\bar{x}_n$ and x_{n+1}.

b) Show that

$$ns_{n+1}^2 = (n-1)s_n^2 + \frac{n}{n+1}\,(x_{n+1} - \bar{x}_n)^2$$

a) $\bar{X}_{n+1} = \dfrac{\sum_{i=1}^{n+1} x_i}{(n+1)} = \dfrac{\sum_{i=1}^{n} x_i + x_{n+1}}{n+1} = \dfrac{\sum_{i=1}^{n} x_i}{n+1} + \dfrac{x_{n+1}}{n+1}$

$= \left(\dfrac{n}{n+1}\right)\left(\dfrac{n+1}{n}\right)\left(\dfrac{\sum_{i=1}^{n} x_i}{n+1} + \dfrac{x_{n+1}}{n+1}\right) = \left(\dfrac{n}{n+1}\right)\left(\dfrac{\sum_{i=1}^{n} x_i}{n} + \dfrac{x_{n+1}}{n}\right)$

$= \boxed{\dfrac{n\bar{x}_n + x_{n+1}}{n+1} = \bar{X}_{n+1}}\ *$ $\Rightarrow$ The equation is meaningful since multiplying mean with previous el. number and adding the new number gives total. Dividing it by new el. number gives new mean.

b) Using the alternative formula of sample variance

$s_n^2 = \dfrac{1}{(n-1)}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2\right)$ and $*$ from part A $\Rightarrow$

$n s_{n+1}^2 - (n-1)s_n^2 = n \cdot \dfrac{1}{n}\left(\sum_{i=1}^{n+1} x_i^2 - (n+1)\bar{x}_{n+1}^2\right) - (n-1)\cdot\dfrac{1}{(n-1)}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2\right)$

$= \sum_{i=1}^{n+1} x_i^2 - (n+1)\bar{x}_{n+1}^2 - \sum_{i=1}^{n} x_i^2 + n\bar{x}_n^2 = x_{n+1}^2 - (n+1)\left(\dfrac{n\bar{x}_n + x_{n+1}}{n+1}\right)^2 + n\bar{x}_n^2$

$= x_{n+1}^2 - \dfrac{n^2\bar{x}_n^2 + x_{n+1}^2 + 2n\bar{x}_n x_{n+1}}{n+1} + n\bar{x}_n^2$

$= \dfrac{(n+1)x_{n+1}^2 - n^2\bar{x}_n^2 - x_{n+1}^2 - 2n\bar{x}_n x_{n+1} + (n^2+n)\bar{x}_n^2}{n+1}$

$= \dfrac{n x_{n+1}^2 - 2n\bar{x}_n x_{n+1} + n\bar{x}_n^2}{n+1} = \dfrac{n}{n+1}\left(x_{n+1}^2 - 2\bar{x}_n x_{n+1} + \bar{x}_n^2\right) = \dfrac{n}{n+1}(x_{n+1} - \bar{x}_n)^2$

$\Rightarrow n s_{n+1}^2 - (n-1)s_n^2 = \dfrac{n}{n+1}(x_{n+1} - \bar{x}_n)^2 \Rightarrow \boxed{n s_{n+1}^2 = (n-1)s_n^2 + \dfrac{n}{n+1}(x_{n+1} - \bar{x}_n)^2}$