1. Your Name: Answer Key

Consider the infinitely many independent, identical experiments of throwing a pair of dice and the outcome of each experiment is the sum of the two numbers. Let $N \geq 1$ be the number of experiments such that the number 5 or 7 first appears as the outcome. Let further X be that number at the N-th experiment, i.e. X can be either a 5 or 7.

a) Find $P(N = n)$ for $n \geq 1$.

b) Find $P(X = 5)$.

c) Prove or disprove that N and X are independent.

[Ross, p. 50 #25; p. 79 eg 4h; p. 104 #76; p. 272, #24]

(a) $P(5) = P((1,4),(2,3),(3,2),(4,1)) = \frac{4}{36} = p_5$

$P(7) = P((1,6),(2,5),\cdots,(6,1)) = \frac{6}{36} = p_7$

$P(\text{no } 5, \text{no } 7) = 1 - p_5 - p_7 = \frac{26}{36}.$

$$P(N=n) = P(\underbrace{\quad\quad\quad}_{\text{no } 5,\ \text{no } 7}^{n-1} \underset{5 \text{ or } 7}{\overset{n}{|}})$$

$$= (1-p_5-p_7)^{n-1}(p_5+p_7) = \left(\frac{26}{36}\right)^{n-1}\left(\frac{10}{36}\right)$$

(b) $$P(X=5) = \sum_{n=1}^{\infty} P(N=n, X=5)$$

$$= \sum_{n=1}^{\infty} P(\underbrace{\quad\quad\quad}_{\text{no } 5,\ \text{no } 7}^{n-1} \underset{5}{\overset{n}{|}})$$

You can use this blank page.

$$= \sum_{n=1}^{\infty} (1-p_5-p_7)^{n-1} p_5 = \left[\sum_{n=1}^{\infty} (1-p_5-p_7)^{n-1}\right] p_5$$

$$= p_5 \frac{1}{1-(1-p_5-p_7)} = \boxed{\frac{p_5}{p_5+p_7}}$$

(c) $P(N=n)\,P(X=5) = (1-p_5-p_7)^{n-1}(p_5+p_7) \times \frac{p_5}{p_5+p_7}$

$$= (1-p_5-p_7)^{n-1} p_5$$

$$||$$

$P(N=n, X=5) = P\left(\underset{n \neq 5,\, n \neq 7}{+ \quad n-1} \quad \overset{n}{\underset{5}{|}}\right)$

$$= (1-p_5-p_7)^{n-1} p_5$$

Hence N & X are independent.

2. Your Name:______________________________

Consider a city in which the male and female drivers occupy α and $1-\alpha$ fractions of the whole city driver population. In any given year, a male and female driver will have an accident with probability p_M and p_F. Assume that the behavior of each driver is independent from year to year.

Now a driver is randomly chosen. Let A_i be the event that this driver will have an accident in the i-th year. Let M be the event that the randomly chosen driver is a male.

a) Suppose $p_M > p_F$. Show that $P(M|A_1) > p(M)$.

b) Suppose $p_M \neq p_F$. Show that $P(A_2|A_1) > p(A_1)$. [Ross p.101, #6]

(a) $P(M) = \alpha$

$$P(M|A_1) = \frac{P(M \cap A_1)}{P(A_1)} = \frac{P(A_1|M)P(M)}{P(A_1)}$$

$$P(A_1) = P(A_1|M)P(M) + P(A_1|F)P(F)$$

$$= \frac{p_M \alpha}{p_M \alpha + p_F(1-\alpha)}$$

$$P(M|A_1) - P(M)$$

$$= \frac{p_M \alpha}{p_M \alpha + p_F(1-\alpha)} - \alpha = \frac{p_M\alpha - p_M\alpha^2 - p_F\alpha(1-\alpha)}{p_M\alpha + p_F(1-\alpha)}$$

$$= \frac{\alpha(1-\alpha)(p_M - p_F)}{p_M\alpha + p_F(1-\alpha)} > 0 \quad (\because p_M > p_F)$$

You can use this blank page.

(b) $P(A_2|A_1) - P(A_1)$

$$= \frac{P(A_2 \cap A_1)}{P(A_1)} - P(A_1)$$

$$= \frac{P(A_2 \cap A_1 | M)P(M) + P(A_2 \cap A_1 | F)P(F)}{P(A_1|M)P(M) + P(A_1|F)P(F)} - P(A_1)$$

$$= \frac{p_M^2 \alpha + p_F^2(1-\alpha) - [p_M \alpha + p_F(1-\alpha)]^2}{P(A_1)}$$

Numerator

$$= p_M^2\alpha + p_F^2(1-\alpha) - p_M^2\alpha^2 - p_F^2(1-\alpha)^2 - 2p_M p_F \alpha(1-\alpha)$$

$$= p_M^2(\alpha-\alpha^2) + p_F^2\underbrace{[(1-\alpha)-(1-\alpha)^2]}_{(1-\alpha)(1-1+\alpha)} - 2p_M p_F\alpha(1-\alpha)$$

$$= \alpha(1-\alpha)[p_M^2 + p_F^2 - 2p_M p_F]$$

$$= \alpha(1-\alpha)(p_M - p_F)^2 > 0 \qquad (p_M \neq p_F)$$

3. Your Name:______________________________

Let $X_1, X_2, \ldots, X_n$ be a collection of iid exponential random variables with parameter λ. Let

$$\begin{aligned} Y_1 &= X_1, \\ Y_2 &= X_1 + X_2, \\ Y_3 &= X_1 + X_2 + X_3, \\ \cdot\cdot &= \cdots\cdots, \\ Y_n &= X_1 + \cdots + X_n. \end{aligned}$$

[Ross, p. 265, example 7e]

Find the joint pdf $p(y_1, y_2, \ldots y_n)$ of $Y_1, Y_2, \ldots, Y_n$.

$$P_Y(y_1, y_2, \ldots, y_n) = \frac{P_X(x_1, x_2, \ldots, x_n)}{\left|\frac{dy_1 \cdots dy_n}{dx_1 \cdots dx_n}\right|}$$

$$\begin{aligned} P_X(x_1, x_2, \ldots, x_n) &= (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2}) \cdots (\lambda e^{-\lambda x_n}) \\ &= \lambda^n e^{-\lambda(x_1 + x_2 + \cdots + x_n)} \\ &= \lambda^n e^{-\lambda y_n} \end{aligned}$$

$$\left|\frac{dy_1 \cdots dy_n}{dx_1 \cdots dx_n}\right| = \det \begin{vmatrix} 1 & 0 & \cdots & 0 \\ \vdots & 1 & & \\ \vdots & \vdots & 1 & 0 \\ & 1 & & \ddots \; 1 \end{vmatrix} = 1$$

Hence $\boxed{P_Y(y_1, \ldots, y_n) = \lambda^n e^{-\lambda y_n}}$

4. Your Name:______________________________

The pdf $p(x)$ of the Gamma distribution with parameter $(\alpha > 0, \lambda > 0)$ is given by:

$$p(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Let X and Y be independent Gamma distributed random variables with parameters (α, λ) and (β, λ). Show analytically that $X+Y$ has a Gamma distribution with parameter $(\alpha+\beta, \lambda)$.

Show how as a by-product that the above conclusion leads to the following integration identity for $\alpha, \beta > 0$:

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

(You are welcome to prove the above identity by any other method.)

$Z = X+Y$

[Ross, p.242, Proposition 3.1]

$$p_Z(z) = \int_0^z p_\alpha(x)\, p_\beta(z-x)\,dx$$

$$= \int_0^z \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\cancel{\Gamma(\alpha)}} \frac{\lambda e^{-\lambda(z-x)}(\lambda(z-x))^{\beta-1}}{\cancel{\Gamma(\beta)}}\,dx$$

$$= \frac{\lambda^{2+\alpha-1+2+\beta-1}\, e^{-\lambda(x+z-x)}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^z x^{\alpha-1}(z-x)^{\beta-1}\,dx$$

Set $w = \frac{x}{z}$

You can use this blank page.

$$= \frac{\lambda^{2+\alpha+\beta} e^{-\lambda z}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 (zw)^{\alpha-1} z^{\beta-1} (1-w)^{\beta-1} z \, dw$$

$$= \frac{\lambda^{2+\alpha+\beta} e^{-\lambda z} z^{\alpha-1+\beta-1+1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 w^{\alpha-1}(1-w)^{\beta-1} dw$$

$$= \left[\frac{\int_0^1 w^{\alpha-1}(1-w)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \right] \lambda e^{-\lambda z} (\lambda z)^{\alpha+\beta-1}$$

a constant. in the form of Gamma($\alpha+\beta, \lambda$)

Hence in order for $f_Z(z)$ to be a pdf, $\int_0^\infty f_Z(z) dz = 1$

thus the constant must be $\frac{1}{\Gamma(\alpha+\beta)}$.

i.e. $\boxed{\int_0^1 w^{\alpha-1}(1-w)^{\beta-1} dw = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$

5. Your Name:______________________________

Let $X_1, \dots, X_n$ be a collection of iid random variables with expectations and variances equal to μ and σ^2. Define the "sample mean" $\overline{X}$ and "sample variance" S^2 as

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n), \quad S^2 = \frac{1}{n}\sum_{i-1}^{n}(X_i - \overline{X})^2$$

Compute $\mathrm{Var}(\overline{X})$ and $\mathrm{E}(S^2)$. [Ross, p. 307, Example 4a]

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2}\left[\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)\right] \quad (\because X_1, X_2, \dots, X_n \text{ independent})$$

$$= \frac{1}{n}\mathrm{Var}(X_1) = \frac{\sigma^2}{n}.$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

~~Method 1~~ $\cancel{E(S^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i - \bar{X})^2}$

$\cancel{= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - 2E}$

You can use this blank page.

Method 1

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\bar{X} + \bar{X}^2\right)$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}X_i^2 - 2\underbrace{\left(\sum_{i=1}^{n}X_i\right)}_{=(n\bar{X})}\bar{X} + \sum_{i=1}^{n}\bar{X}^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}X_i^2 - 2n(\bar{X})^2 + n(\bar{X})^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}X_i^2 - n(\bar{X})^2\right] \qquad \left(\mathrm{Var}(X) = EX^2 - (EX)^2\right)$$

$$\mathrm{Var}(S^2) = \frac{1}{n}\left[n\left(\mathrm{Var}(X_1)^2 + \mu^2\right) - n\,E\left[(\bar{X})^2\right]\right]$$

$$= \sigma^2 + \mu^2 - E\left[(\bar{X})^2\right] \qquad \left(E(\bar{X}) = \mu\right)$$

$$= \sigma^2 + \mu^2 - \mathrm{Var}(\bar{X}) - (E\bar{X})^2$$

$$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \sigma^2 - \frac{\sigma^2}{n} = \boxed{\left(\frac{n-1}{n}\right)\sigma^2}$$

You can use this blank page.

$\boxed{\text{Method 2}}$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{X_1 + \cdots + X_n}{n}\right)^2$$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\left(\underline{nX_i - (X_1 + \cdots + X_n)}\right)^2$$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\left((n-1)X_i - \sum_{j\neq i}X_j\right)^2$$

n-1 so many

$$= \frac{1}{n^3}\sum_{i=1}^{n}\left[(n-1)(X_i-\mu) - \sum_{j\neq i}(X_j-\mu)\right]^2$$

Hence

$$ES^2 = \frac{1}{n^3}\sum_{i=1}^{n}E\left(\left[(n-1)(X_i-\mu) - \sum_{j\neq i}(X_j-\mu)\right]^2\right)$$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\left\{(n-1)^2 E\left((X_i-\mu)^2\right) + \sum_{j\neq i}E\left[(X_j-\mu)^2\right]\right.$$

$$\left. + E(\text{cross terms}) \rightarrow 0\right\}$$

You can use this blank page.

(The cross ~~ ~~ terms are zero as X_i, X_j are ind.)

$$= \frac{1}{n^3} n \left\{ (n-1)^2 \sigma^2 + (n-1) \sigma^2 \right\}$$

$$= \frac{\cancel{n}(n-1)(\cancel{n-1+1})}{n^{\cancel{3}}} \sigma^2$$

$$= \boxed{\left(\frac{n-1}{n}\right) \sigma^2}$$

6. Your Name:______________________________

(Estimation of the length of an interval.) Let $L > 0$ be some unknown but fixed length. Let $X_1, X_2, \ldots$ be a sequence of iid random variables uniformly distributed on $[0, L]$. The goal is to use the X_i's to estimate L.

a) Let $A_n = 2\dfrac{X_1 + \cdots + X_n}{n}$. Show that A_n is an unbiased estimator in the sense that $\mathrm{E}(A_n) = L$.

b) Let $B_n = \gamma_n \max\{X_1, X_2, \ldots, X_n\}$ where γ_n is some number. Find the correct value of γ_n such that B_n is also an unbiased estimator, i.e. $\mathrm{E}(B_n) = L$.

(Hint: find the distribution of B_n first.)

c) Find $\mathrm{Var}(A_n)$ and $\mathrm{Var}(B_n)$.

d) Which estimator is "more superior"?

[Breiman: Statistics, view toward Applications, p.55]

(a) $$A_n = 2\left(\frac{X_1 + \cdots + X_n}{n}\right)$$

$$EA_n = \frac{2}{n}\, n\, E(X_1) = \frac{2}{n}\, n\, \frac{L}{2} = L$$

(b) $$B_n = \gamma_n \underbrace{\max(X_1, \ldots, X_n)}_{M_n}$$

$$P(M_n \leq t) = P(X_1 \leq t)P(X_2 \leq t) \cdots P(X_n \leq t)$$
$$= \left(\frac{t}{L}\right)^n$$

$$P(t) = n t^{n-1} / L^n$$

You can use this blank page.

$$E(B_n) = \gamma_n E(M_n) = \gamma_n \int_0^L t \frac{n t^{n-1}}{L^n} dt$$

$$= \frac{n\gamma_n}{L^n} \int_0^L t^n dt = \frac{n}{n+1} \gamma_n L = L$$

Hence $\gamma_n = \frac{n+1}{n}$

(c) $Var(A_n) = Var\left(\frac{X_1 + \cdots + X_n}{n} \times 2\right)$

$$= \frac{2^2}{n^2} n \, Var(X_1) = \frac{4}{n} \frac{L^2}{12} = \frac{L^2}{3n}$$

$$Var(B_n) = Var(\gamma_n M_n) = \gamma_n^2 Var(M_n)$$

$$= \gamma_n^2 \left[E(M_n^2) - (EM_n)^2 \right]$$

You can use this blank page.

$$E(M_n^2) = \int_0^L t^2 \frac{n t^{n-1}}{L^n} dt = \frac{n}{L^n} \int_0^L t^{n+1} dt$$

$$= \left(\frac{n}{n+2}\right) L^2$$

$$(E M_n)^2 = \left(\frac{n}{n+1}\right)^2 L^2$$

$$\mathrm{Var}(B_n) = \gamma_n^2 \left[\left(\frac{n}{n+2}\right) - \frac{n^2}{(n+1)^2}\right] L^2$$

$$= L^2 \gamma_n^2 \left[\frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\right]$$

$$= L^2 \frac{\cancel{(n+1)^2}}{n^2} \left[\frac{\cancel{n^3} + \cancel{2n^2} + n - \cancel{n^3} - \cancel{2n^2}}{(n+2)\cancel{(n+1)^2}}\right]$$

$$= \frac{L^2}{n(n+2)}$$

You can use this blank page.

1 d) $Var(A_n) = \frac{L^2}{3n} = O\left(\frac{1}{n}\right) \xrightarrow{n\to\infty} 0$

$Var(B_n) = \frac{L^2}{n(n+2)} = O\left(\frac{1}{n^2}\right) \xrightarrow{n\to\infty} 0$

~~Both~~ Both $Var(A_n)$ & $Var(B_n)$ go to zero as $n \to \infty$ but $Var(B_n)$ goes to zero at a faster rate $\left(O\left(\frac{1}{n^2}\right) \ll O\left(\frac{1}{n}\right)\right)$

Hence Bn is more superior.

(Variance measures fluctuations.).