Chapter 1 - Probability - Exercises

1) Prove the theorem of continuity of Probabilities: If $A_n \to A$ then $P(A_n) \underset{n\to\infty}{\to} P(A)$

Let $i < j$. Since $B_i \subset A_i$ and $B_j \cap A_i = \emptyset$ $\Rightarrow$ B_i and B_j : disjoint

Since $A_1 \subset A_2 \subset \dots$ it follows that $A_n = \bigcup_{i=1}^{n} A_i$ for each n. Suppose $\bigcup_{i=1}^{n} B_i = A_n$ for some n. It follows $\bigcup_{i=1}^{n+1} B_i = A_n \cup B_{n+1} = (\bigcup_{i=1}^{n} A_i) \cup (A_{n+1} \setminus (\bigcup_{i=1}^{n} A_i)) = \bigcup_{i=1}^{n+1} A_i$

Let $A_1 \supset A_2 \supset \dots$ be monotonous decreasing. Then $A_1^c \subset A_2^c \subset \dots$ $\Rightarrow$ monotonous increasing

$P(\bigcap_n A_n) = 1 - P(\bigcup A_n^c) = 1 - \lim_n P(A_n^c) = \lim_n P(A_n)$

2) Prove $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Since $P(\emptyset \cup \emptyset) = 2 \cdot P(\emptyset)$ By additivity $\Rightarrow P(\emptyset) = 0$. If A contained in B:

$P(B) = P(A \cup B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A)$

$\Rightarrow P(A) \leq P(\Omega) = 1$

since $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1 \Leftrightarrow P(A) = 1 - P(A^c)$

Taking $A_2 = A_3 = \dots = \emptyset$ in the countable additivity property (Axiom 3) we obtain $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ for any disjoint sets A_1, A_2.

5) Toss a coin until we get two heads. What is the sample space S? What's the prob that exactly k tosses are required?

$S = \{H, T\}^{\infty}$, Let X_n : be 1 when n-toss is heads and zero otherwise (binomial)

$P(X_1 + \dots + X_{k-1} = 1) \cdot P(X_k = 1) = \binom{k-1}{1} \cdot p(1-p)^{k-2} \cdot p = (k-1) p^2 (1-p)^{k-2}$

if fair coin then it simplifies to $(k-1) \cdot 2^{-k}$.

6) Let $\Omega = \{0, 1, \dots\}$. Prove that there doesn't exist a uniform distribution on Ω

~~Let P~~ By additivity $1 = P(N) = \sum_n P(\{n\})$. Suppose P is uniform. Then $P(\{n\}) = c$ for ~~any~~ each n and hence $P(N) = c \cdot \infty$ ($0 \cdot \infty = 0 \to$ contradiction).

10) Monty Hall Problem. A prize is placed behind one of three doors. Let say you pick Door 1 and Monty ... of the remaining and it is empty and ask you if you want to change your door. Should you stay or switch? Correct answer is that you should switch. Prove it.

The player is between door 1 and door 3:

$p_i = P(W_1 = i | W_2 = 2) = \frac{P(W_2 = 2 | W_1 = i)\, P(W_1 = i)}{P(W_2 = 2)} = \frac{P(W_2 = 2 | W_1 = i)}{3 \cdot P(W_2 = 2)}$

$P(W_2 = 2 | W_1 = i) = \begin{cases} 1/2 & \text{if } i = 1 \\ 1 & \text{if } i = 3 \end{cases}$

11) A, B independent, show A^c & B^c independent events

$P(A^c \cap B^c) = P(A^c) - P(B) + P(A \cap B)$ using the independence of A & B

$= P(A^c) - P(B) + P(A)P(B) = P(A^c) - (1-P(A))P(B) = P(A^c) - P(A^c)P(B)$

$= P(A^c)[1-P(B)] = P(A^c) \cdot P(B^c)$

13) Suppose fair coin and tossed repeatedly until both a head & tail appeared at least once. Find Ω and what is the prob that we need 3 tosses?

$S = \{H, T\}$ We stop at the 3rd toss if and only if: HHT or TTH, let p = prob of heads then $p^2(1-p) + (1-p)^2 p = p(1-p) = \frac{1}{4}$

? → 15) Prob of a child to have blue eyes = 1/4. Assume independence between children. Family has 3 children. a) if it known that at least one child has blue eyes, what is the prob that at least two children have blue eyes? b) If it is known that the youngest child has blue eyes, what is the probability that at least 2 have blue eyes?

Let B_k be RV that and only if the k-th child has blue eyes. Let $B = B_1 + B_2 + B_3$. Let $p = 1/4$ that having blue eyes, $\frac{\frac{9}{64} + \frac{1}{64}}{1 - \frac{27}{64}} = \frac{10}{37}$ Exactly 2 blue eyes $\binom{3}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \frac{3}{4} = \frac{9}{64}$

a) $P(B \geq 2 \mid B \geq 1) = \dfrac{P(B \geq 2)}{P(B \geq 1)} = \dfrac{1 - P(B \leq 1)}{1 - P(B=0)}$ $\quad P(B=0) = q^3 \quad P(B=1) = 3pq^2$

Exactly 3 blue eyes: $\left(\frac{1}{4}\right)^3 = \frac{1}{64}$ → no blue eyes $= \left(\frac{3}{4}\right)^3 = \frac{27}{64}$

$\Rightarrow P(B \geq 2 \mid B \geq 1) = \dfrac{1 - q^3 - 3pq^2}{1 - q^3} = \dfrac{10}{37}$

b) $P(B \geq 2 \mid B_1 = 1) = \dfrac{P(B_1 = 1, B_2 + B_3 \geq 1)}{P(B_1 = 1)} = P(B_2 + B_3 \geq 1) = 1 - P(B_2 + B_3 = 0)$

$= 1 - q^2 = 7/16 \quad 1 - \frac{9}{16} = \frac{7}{16}$ both haven't blue eyes $\left(\frac{3}{4}\right)^2 = \frac{9}{16}$ one of the two has blue eyes $\binom{2}{1}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{6}{16}$

19) 30% computer owners use Macintosh, 50% windows and 20% Linux. Suppose 65% of Mac took a virus, 82% of windows and 50% of Linux. We select a random person and learn that she was infected. What the prob to use windows?

$P(W \mid V) = \dfrac{P(V \mid W) \cdot P(W)}{P(V)} = \dfrac{P(V \mid W) \cdot P(W)}{\sum_{x \in \{M, W, L\}} P(V \mid x) \cdot P(x)} = \dfrac{82 \times 50}{65 \cdot 30 + 82 \cdot 50 + 50 \cdot 20} = \dfrac{82}{141} = 58\%$

Chapter 2: Random Variables - Exercises

1) prove $P(X = x) = F(x^+) - F(x^-)$ using lemma 2.15

$P(X = x) = F(x^+) - F(x^-)$ since F is right continuous $F(x) = F(x^+)$

2) Let X: $P(X=2) = P(X=3) = 1/10$ and $P(X=5) = 8/10$ Plot CDF F. Find $P(2 < X \leq 4.8)$ and $P(2 \leq X \leq 4.8)$

By lemma 2.15: $P(2 < X \leq 4.8) = F(4.8) - F(2) = 1/10$

$P(2 \leq X \leq 4.8) = P(X=2) + P(2 < X \leq 4.8) = F(4.8) - F(2^-) = 2/10$

3) Prove lemma 2.15: Since F monotone we can write $F(x^-) = \lim F(x_n)$ where (x_n) strictly increasing sequence converging to x. Let $A_n = \{X \leq x_n\}$ so that $\{X < x\} = \bigcup_n A_n$. By the continuity of probability $P(X < x) = \lim_n P(A_n) = \lim_n F(x_n)$

$$F_X(x) = \begin{cases} 0, & x<0 \\ \frac{x}{4}, & 0 \le x < 1 \\ \frac{1}{4}, & 1 \le x < 3 \\ \frac{1}{4} + \frac{3(x-3)}{8}, & 3 \le x \le 5 \\ 1, & x \ge 5 \end{cases}$$ use $F(x) = \int_{-\infty}^{x} f(t)\,dt$, eg. $3 \le x < 5$: $F(x) = \frac{1}{4} + \int_3^x \frac{3}{8}\,dt = \dots$

4) X with probability density function $f_X(x) = \begin{cases} 1/4, & 0 < x < 1 \\ 3/8, & 3 < x < 5 \\ 0, & \text{otherwise} \end{cases}$

a) Find the CDF b) if $Y = 1/X$. Find the PDF of $f_Y(y)$ for Y

a) $F_X(x) = \frac{x}{4} I_{[0,1)}(x) + \frac{1}{4} I_{[1,\infty)}(x) + \frac{3}{8}(x-3) I_{[3,5]}(x) + \frac{3}{4} \cdot I_{[5,+\infty)}(x)$

b) $F_X(0) = 0$ and $Y = 1/X \Rightarrow F_Y(0) = 0$ for $y > 0$

$F_Y(y) = P(X \geq 1/y) = 1 - P(X < 1/y) = 1 - F_X(1/y)$

5) X, Y: discrete RVs. Show X, Y: independent if and only if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all x, y

suppose X, Y: independent: then $f_{X,Y}(x,y) = P(X \in \{x\}, Y \in \{y\}) = P(X \in \{x\}) P(Y \in \{y\}) = f_X(x) f_Y(y)$. To establish the converse suppose that $f_{XY} = f_X \cdot f_Y$. For a subset A of the support of X and B of the support of Y

$P(X \in A, Y \in B) = \sum_{(x,y) \in A \times B} f_{XY}(x,y) = \sum_{x \in A} f_X(x) \cdot \sum_{y \in B} f_Y(y) = P(X \in A) P(Y \in B)$

6) Let $I_A(x)$ be the indicator function of A.

$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$ Let $Y = I_A(X)$ find an expression for the CDF

$F_Y(y) = \begin{cases} 0, & \text{if } y < 0 \\ P(X \notin A), & \text{if } 0 \le y < 1 \\ 1, & \text{if } y \geq 1 \end{cases}$

7) X, Y independent $\sim$ Uniform$(0,1)$. Let $Z = \min\{X, Y\}$. Find the density $f_Z(z)$ for Z.

Since $P(Z > z) = P(\min\{X,Y\} > z) = P(X > z) P(Y > z) = (1 - F_X(z))(1 - F_Y(z))$

$\Rightarrow F_Z(z) = 1 - (1 - F_X(z))(1 - F_Y(z)) = F_X(z) + F_Y(z) - F_X(z) \cdot F_Y(z)$

when X, Y: same distribution F, $F_Z(z) = 2F(z) - F(z)^2$

$\Rightarrow f_Z(z) = 2f(z) - 2F(z) \cdot f(z)$. When $F \sim U(0,1)$: $f_Z(z) = 2(1-z) \cdot I_{(0,1)}(z)$

8) Let X have CDF F. Find CDF of $X^+ = \max\{0, X\}$.

Let $Y = X^+$. Note $F_Y(0^-) = 0$ and $F_Y(0) = F_X(0) \Rightarrow F_Y(x) = F_X(x)$ for $x > 0$

10) X, Y independent. Show $g(X)$ independent of $h(Y)$, g, h: functions

$P(g(X) \in A, h(Y) \in B) = P(X \in g^{-1}(A), Y \in h^{-1}(B))$

$= P(X \in g^{-1}(A)) \, P(Y \in h^{-1}(B)) = P(g(X) \in A) \cdot P(h(Y) \in B)$.

11) Toss a coin once and p = prob of heads. X: "number of heads" ($X = 0$ or 1), Y = "number of tails" ($Y = 0$ or 1)

a) Prove X, Y independent. b) $N \sim$ Poisson(λ), we toss a coin N times. let $X + Y$ number of heads & tails. Show X, Y independent. $X + Y = 1$ (one toss)

a) $P(X = 1, Y = 0) = 0 \neq p(1-p) = P(X=1) \cdot (Y=0)$. $\Rightarrow$ dependent.

b) $P(X = i, Y = j) = \frac{\lambda^{i+j} e^{-\lambda}}{(i+j)!} \binom{i+j}{i} p^i (1-p)^j = e^{-\lambda} \frac{\lambda^i p^i}{i!} \frac{\lambda^j (1-p)^j}{j!}$ $\Rightarrow$ independent

$P(X=1) = p$, $P(Y=1) = 1-p$, $P(X=0) = 1-p$, $P(Y=0) = p$; $P(X=1 \,\&\, Y=0) = p$, $P(X=1) P(Y=0) = p \cdot p = p^2$, $p \neq p^2 \Rightarrow$ dependent

12) Prove. Suppose that the range of X and Y is a (possibly infinite) rectangle

If $P(x,y) = g(x)\, h(y)$ for some functions g & h (not necessarily density functions) then X and Y are independent.

If X,Y joint density satisfying $f(x,y) = g(x)\cdot h(y)$ then

$P(X\le x, Y\le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,dt\,ds = \int_{-\infty}^{x} g(s)\,ds \int_{-\infty}^{y} h(t)\,dt$

marginal distr. of $X \Rightarrow P(X\le x) = C_h \int_{-\infty}^{x} g(s)\,ds$ where $C_h = \int_{-\infty}^{\infty} h(t)\,dt$, it follows that $f_X = h\, C_h$ for C_g $f_Y = g\, C_g$, $C_g C_h = 1 \Rightarrow C_g = \frac{1}{C_h}$

$\Rightarrow f_{X,Y} = f_X f_Y$

since area of a unit disc is π, the joint PDF $f_{XY}(x,y) = \frac{1}{\pi}$, for $x^2+y^2\le 1$

14) (X,Y) uniformly distributed on the unit disk $\{(x,y)\} : x^2+y^2 \le 1\}$

Let $R = \sqrt{X^2+Y^2}$ Find CDF and PDF in R $F_R(r) = P(R\le r) = P(\sqrt{X^2+Y^2}\le r) = P(X^2+Y^2\le r^2)$

Let $0<r<1$ then $F_R(r) = \pi r^2/\pi = r^2$ hence $f_R(r) = 2r$

17) Let $f_{XY}(x,y) = \begin{cases} c(x+y)^2, & 0\le x\le 1 \quad 0\le y\le 1 \\ 0 & \text{otherwise}\end{cases}$

Find $P(X<\frac{1}{2} \mid Y=\frac{1}{2})$

$f_Y(1/2) = \int_0^1 f(x,1/2)\,dx = c\int_0^1 (x+\frac{1}{4})dx = \frac{3}{4}c$

$f_{X|Y}(x|y) = \frac{f_{XY}(x,1/2)}{f_Y(1/2)} = \frac{4}{3}(x+\frac{1}{4})\cdot I_{(0,1)}(x)$

$\Rightarrow P(X<1/2 \mid Y=1/2) = \frac{4}{3}\int_0^{1/2}(x+\frac{1}{4})dx = \frac{1}{3}$

Chapter 3: Expectations Exercises

1) Play a game and start with c dollars. On each play either you double or halve your money (equal prob). What is the expected fortune after n trials

Let X_n: number of dollars at the n-th trial. Then:

$E[X_{n+1}|X_n] = \frac{1}{2}(2X_n + \frac{1}{2}X_n) = \frac{5}{4}X_n$ By the rule of iterated expectations

$EX_{n+1} = 5/4\, E X_n$ By induction $\Rightarrow EX_n = (5/4)^n c$

2) Show $V(X)=0$ if and only if there is a constant c such that $P(X=c)=1$

If $P(X=c)=1$ then $E(X^2) = (EX)^2 = c^2 \Rightarrow V(X)=0$

whenever the Y nonnegative RV, $EY=0 \Rightarrow P(Y=0)=1$ in this case

$Y=(X-EX)^2 \Rightarrow P(X=EX)=1$

suppose $EY=0$. Take $A_n=\{Y\ge 1/n\}$ then $0=EY=E[YI_{A_n}+YI_{A_n^c}] \ge E[YI_{A_n}] \ge \frac{1}{n}P(A_n)$

$P(A_n)=0$ for all n. By continuity of prob. $P(Y>0) = P(\cup_n A_n) = \lim P(A_n) = 0$

3) Let $X_1,\dots,X_n \sim \text{Uniform}(0,1)$, $Y_n = \max\{X_1,\dots,X_n\}$ Find $E(Y_n)$.

Since $F_{Y_n}(y) = P(X_1\le y)^n = y^n \Rightarrow f_{Y_n}(y) = ny^{n-1} \Rightarrow EY_n = n\int_0^1 y\,y^{n-1}dy = \frac{n}{n+1}$

4) Random walk $\rightarrow$ Find $E(X_n)$, $V(X_n)$

$X_n = \sum_{i=1}^{n} (1-2B_i) = n - 2\sum_i B_i$ where $B_1 \dots B_n \sim$ Bernoulli(p) are IID

It follows that $EX_n = n - 2n\,EB_1 = n - 2np$ and $V(X_n) = 4n\,V(B_1) = 4np(1-p)$

5) Fair coin is tossed until head is obtained. What is the expected number of tosses required?

Geometric Distribution: Let T = number of tosses until a heads is observed. Let C denote the result of the first toss. $E_T = \frac{1}{2}(E[T|C=H] + E[T|C=T]) = \frac{1}{2}(1 + (1+E_T))$ $P(\text{getting a head}) = \frac{1}{2}$

$\Rightarrow E_T = 2$ $\qquad E[X]$ of geometric $= \frac{1}{p} = \frac{1}{0.5} = 2$

*10) Let $X \sim N(0,1)$ $\quad Y = e^X$ $\quad$ Find $E(Y)$ and $V(Y)$

The MGF of a normal random variable is $\exp(t^2/2)$. Therefore $E\exp(X) = \sqrt{e}$ and

$V(\exp(X)) = E[\exp(2X)] - (E\exp(X))^2 = e^2 - e$

13) We generate a RV. First flip a coin. If heads take $X \sim$ Unif$(0,1)$. If tails take $\sim$ Unif$(3,4)$. a) Find the mean of X. b) Find σ of X.

C. result of the coin toss

$EX = E[\text{Unif}(0,1)\,I_{(C=H)} + \text{Unif}(3,4)\,I_{(C=T)}] = \frac{1}{2}(E\,\text{Unif}(0,1) + E\,\text{Unif}(3,4)) = 2$

$E[X^2] = \frac{1}{2}(E[\text{Unif}(0,1)^2] + E[\text{Unif}(3,4)^2]) = 19/3$ $\quad V(X) = 19/3 - 4 = 7/3$ $\quad sd(X) = \frac{\sqrt{21}}{3}$

15) Let $f_{X,Y}(x,y) = \begin{cases} 1/3\,(x+y) & 0 \le x \le 1 \quad 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$

Find $V(2X - 3Y + 8)$

$E[(2X-3Y)^2] = \int_0^2 \int_0^1 (2x-3y)^2 \cdot \frac{1}{3}(x+y)\,dx\,dy = 86/9$

$E[2X-3Y] = \int_0^2 \int_0^1 (2x-3y)\frac{1}{3}(x+y)\,dx\,dy = -\frac{23}{9}$

$V(2X-3Y) = 245/81$

22) Let $X \sim$ Uniform$(0,1)$. Let $0 < a < b < 1$. Let $Y = \begin{cases} 1 & 0 < x < b \\ 0 & \text{otherwise} \end{cases}$

$Z = \begin{cases} 1 & a < x < 1 \\ 0 & \text{otherwise} \end{cases}$ a) Are Y & Z independent?

b) Find $E(Y|Z)$

a) Note that $E[YZ] = E\,I_{(a,b)}(X) = b-a$. Moreover $EY = E\,I_{(0,b)}(X) = b$ and

$EZ = E\,I_{(a,1)}(X) = 1-a$. Since $E(YZ) \neq EY\,EZ \Rightarrow$ dependent

b) If $Z=0$, then $X \le a < b \Rightarrow Y = 1$. Therefore $E[Y|Z=0] = 1$

$E[Y|Z=1] = \frac{E[YZ]}{P(Z=1)} = \frac{b-a}{1-a}$

Chapter 4. Inequalities - Exercises

1) Let $X \sim$ Exponential(β). Find $P(|X - \mu_X| \ge k\sigma_X)$ for $k > 1$. Compare this to the bound you get from Chebyshev's inequality. $\le \frac{\sigma^2}{k^2\sigma^2}$

Chebyshev's inequality $P(|X-\mu| \ge k\sigma) \le 1/k^2$. An exact calculation yields instead $e^{-(1+k)}$. To see this note that $\beta(\mu \pm k\sigma) = 1 \pm k$ and $1-k < 0$ so that

$P(|X-\mu| \le k\sigma) = P(X \le \mu + k\sigma) = F(\mu + k\sigma) = 1 - e^{-1-k}$

2) Let $X \sim$ Poisson(λ). Use Chebyshev's inequality to show that $P(X \geq 2\lambda)$

$$\mathbb{P}(X \geq 2\lambda) = P(X-\lambda \geq \lambda) = P(|X-\lambda| \geq \lambda) \leq \frac{1}{\lambda}$$

4) Let $X_1 \ldots X_n \sim$ Bernoulli(p)

a) Let $a > 0$ be fixed & defined. $\varepsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{a}\right)}$ Let $\hat{p_n} = n^{-1} \sum_{i=1}^{n} X_i$

Define $C_n = (\hat{p_n} - \varepsilon_n \quad \hat{p_n} + \varepsilon_n)$. Use Hoeffding's inequality to show

$\mathbb{P}(C_n$ contains $p) \geq 1-a$

$$\mathbb{P}(p \in C_n) = 1 - P(p \notin C_n) \geq 1 - 2\exp(-2n \cdot \varepsilon_n^2) = 1 - 2\exp\left(\log\left(\frac{a}{2}\right)\right) = 1-a$$

c) Plot the length of the interval versus n. Suppose we want the length of the interval to be no more than 0,05. How large should n be?

The length of the interval is $2 \cdot \varepsilon_n$. This length is at most $c > 0$, if and only if $n \geq 2\log(2/a)/c^2$

6) Let $Z \sim N(0,1)$. Find $\mathbb{P}(|Z| > t)$ and plot this as a function of t. From Markov's inequality we have the bound $P(|Z| > t) \leq \frac{E|Z|^k}{t^k}$ for any $k > 0$. Plot these bounds for $k = 1, 2, 3, 4, 5$ and compare them to the true value of $P(|Z| > t$. Also plot the bound from Mill's inequality.

Chapter 5: Convergence of RV: - Exercises

1) Let $X_1 \ldots X_n$ be IID with finite mean $\mu = E(X_1)$ and finite variance $\sigma^2 = V(X_1)$. Let $\bar{X_n}$ be the sample mean and let S_n^2 be the sample variance. a) Show that $E(S_n^2) = \sigma^2$, b) Show that $S_n^2 \xrightarrow{P} \sigma^2$.

a) $E\bar{X} = \frac{1}{n} \sum_i EX_i = EX_1 = \mu \qquad V(\bar{X}) = \frac{1}{n^2} \sum_i V(X_1) = \frac{1}{n} V(X_1) = \frac{\sigma^2}{n}$

$$(n-1) \cdot S_n^2 = \sum_i (X_i - \bar{X})^2 = \sum_i X_i^2 - 2 \cdot X_i \bar{X} + \bar{X}^2$$

$$= \frac{n-1}{n} \mathbb{E}[S_n^2] = \mathbb{E}[X_1^2] - 2\mathbb{E}[X_1 \bar{X}] + \mathbb{E}[\bar{X}^2]$$

note that: $\mathbb{E}[X_1^2] = \sigma^2 + \mu^2$ and $\mathbb{E}[\bar{X}^2] = \sigma^2/n + \mu^2$

$$X_1 \bar{X} = \frac{1}{n}\left(X_1^2 + X_1 \cdot \sum_{j \neq 1} X_j\right) \text{ and hence } \mathbb{E}[X_1 \bar{X}] = \sigma^2/n + \mu^2$$

(substitute in above equation): $\mathbb{E}[S_n^2) = \sigma^2$

b) Note that: $S_n^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_i (X_i^2 - 2X_i\bar{X} + \bar{X}^2)$

$$= \frac{1}{n-1} \sum_i X_i^2 - \frac{n}{n-1} \bar{X}^2 = c_n \cdot \frac{1}{n} \sum_i X_i^2 - d_n \bar{X}^2 \text{ where}$$

$c_n \to 1$ and $d_n \to 1$. By the WLLN $\frac{1}{n} \cdot \sum X_i^2$ and $\bar{X}^2$ converge, in probability, to $E[X_1^2]$ and μ^2. By theorem 5.5 (d), $c_n \frac{1}{n} \cdot \sum_i X_i^2$ and $d\bar{X}^2$ converge, in probability, to the same quantities. Lastly by theorem 5.5 (a) S_n^2 converges, in probability, to $E[X_1^2] - \mu^2 = \sigma^2$.

2) Let $X_1, X_2 \ldots$ be a sequence of RV Show that $X_n \xrightarrow{qm} b$ if only if $\lim_{n\to\infty} E(X_n) = b$ and $\lim_{n\to\infty} V(X_n) = 0$

Suppose X_n converges to b in quadratic mean. By Jensen's Inequality $E[X_n - b]^2 \leq E[|X_n - b|]^2 \leq E[(X_n-b)^2] \to 0$ therefore $EX_n \to b$ Next, note that $E[(X_n-b)^2] = E[X_n^2] - 2bE[X_n] + b^2 = V(X_n) + E[X_n]^2 - 2bE[X_n] + b^2$

Taking limits of both sides reveals $\lim_n V(X_n) = 0$ As for the converse, we can apply the limits $\lim_n E[X_n] = b$ and $\lim_n V(X_n) = 0$ directly to the equation above

4) Let $X_1, X_2 \ldots$ be a sequence of RV such that

$P(X_n = \frac{1}{n}) = 1 - \frac{1}{n^2}$ and $P(X_n = n) = \frac{1}{n^2}$ Does X_n converge in probability?

Does X_n converge in quadratic mean?

Let $\varepsilon > 0$. For n sufficiently large.

$P(|X_n - 0| > \varepsilon) = P(X_n > \varepsilon) = P(X_n = n) = 1/n^2 \to 0$ and hence X_n converges in probability However

? $E[(X_n - 0)^2] = E[X_n^2] \geq E[X_n^2 I_{\{X_n = n\}}] = n^2 P(X_n = n) = 1$ and hence X_n does not converge in quadratic mean.

8) Suppose we have a computer program consisting of $n = 100$ pages of code

? Let X_i be the number of errors on the i^{th} page of code. Suppose that the X_i's are Poisson with mean 1 and they're independent. Let $Y = \sum_{i=1}^{n} X_i$ be the total number of errors. Use the CLT to approximate $P(Y < 90)$

Let $\varepsilon > 0$. Then, $P(|X_n - X| > \varepsilon) \leq P(X_n \neq X) = \frac{1}{n} \to 0$ Therefore X_n converges in probability (and hence in distribution) to X. On the other hand $E[(X - X_n)^2] = E[(X - e^n)^2 I_{\{X_n \neq X\}}] = E[1 - 2Xe^n + e^{2n}] P(X_n \neq X)$

$= \frac{1 + e^{2n}}{n} \to \infty$

? 12) Let $X, X_1, X_2, X_3 \ldots$ be RV that are integer valued and positive. Show that $X_n \rightsquigarrow X$ if and only if $\lim_{n\to\infty} P(X_n = k) = P(X = k)$ for every integer k.

Let F be the CDF of an integer valued RV K. Let k be an integer. It follows that $F(k) = F(k + c)$ for all $0 \leq c < 1$. we use this multiple times below.

To prove the forward direction, suppose $X_n \rightsquigarrow X$. By definition $F_{X_n} \to F_X$ at all points of continuity of F_X. Therefore:

$P(X_n = k) = F_{X_n}(k + 1/2) - F_{X_n}(k - 1/2) \to F_X(k + 1/2) - F_X(k - 1/2) = P(X = k)$

To prove the reverse direction suppose $P(X_n = k) \to P(X = k)$ for all integers k. Let j be an integer and note that $F_{X_n}(j) = \sum_{k \leq j} P(X_n = k) \to \sum_{k \leq j} P(X = k) = F_X(j)$

and hence $X_n \rightsquigarrow X$.

1) Find the mean & variance of an exponential distribution: $\mu = \frac{1}{\theta}$, $\sigma^2 = \frac{1}{\theta^2}$

we need to find $P(|X-\mu| \geq k\sigma) = P(|X - \frac{1}{\theta}| \geq k \frac{1}{\theta})$

this can be written also as $P(X \notin [\frac{1}{\theta} - k\cdot\frac{1}{\theta}, \frac{1}{\theta} + k\frac{1}{\theta}]) = P(X \notin [\frac{1-k}{\theta}, \frac{1+k}{\theta}])$

we need to calculate this: $P(X < \frac{1-k}{\theta}) + P(X > \frac{1+k}{\theta})$

The CDF of $X \sim \text{Exponential}(\theta)$ is $F_X(x) = 1 - \cancel{\theta} e^{-\theta x}$ for $x \geq 0$

- $P(X < \frac{1-k}{\theta})$. For $k>1$ this prob is zero since $\frac{1-k}{\theta} < 0$ and X nonnegative
- $P(X > \frac{1+k}{\theta})$: $P(X > \frac{1+k}{\theta}) = 1 - F_X(\frac{1+k}{\theta}) = e^{-\theta\frac{1+k}{\theta}} = e^{-1-k}$

The total prob: $P(|X-\mu| \geq k\sigma) = P(X > \frac{1+k}{\theta}) = e^{-(1+k)}$

Let's compare it with Chebyshev's inequality:

$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$ we know $\mu = \frac{1}{\theta}$, $\sigma = \frac{1}{\theta}$

Exact result is $P(|X-\mu| \geq k\sigma) = e^{-(1+k)}$ as $k\uparrow \Rightarrow$ decays exponentially

The bound from Chebyshev is $\frac{1}{k^2}$ as $k\uparrow \Rightarrow$ decays quadratically.

for large k the Chebyshev bound is less tight as exponential decay is faster than quadratic

2) Poisson (λ) : $\mu = \lambda$, $\sigma = \lambda$

$P(X \geq 2\lambda)$ write as deviation from the mean: $P(X - \lambda \geq \lambda)$

Chebyshev's inequality: $P(|X-\lambda| \geq \lambda) \leq \frac{Var(X)}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$

$\Rightarrow P(|X-\lambda| \geq \lambda) \leq \frac{1}{\lambda}$

4) a) Since $X_1 \dots X_n \overset{iid}{\sim} \text{Bernoulli}(p)$: $P(X_i = 1) = p$, $P(X_i = 0) = 1-p$

sample mean $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (estimator for the true p)

for $a > 0$: $I_n = [\hat{p}_n - \varepsilon_n, \hat{p}_n + \varepsilon_n]$ where $\varepsilon_n = \sqrt{\frac{\log(\frac{2}{a})}{2n}}$, the I_n contains the true p with probability at least $1-a$. We need to show: $P(p \in I_n) \geq 1-a$

Hoeffding's Inequality for Bernoulli random variables (bounded between 0 and 1):

$P(|\hat{p}_n - p| \geq \varepsilon) \leq 2\exp(-2n\cdot\varepsilon^2)$ substitute ε_n

$P(|\hat{p}_n - p| \geq \sqrt{\frac{\log(\frac{2}{a})}{2n}}) \leq 2\exp\left(-2n\left(\sqrt{\frac{\log(\frac{2}{a})}{2n}}\right)^2\right) = 2\exp(-\log(\frac{2}{a}))$

$\Leftrightarrow P(|\hat{p}_n - p| \geq \sqrt{\frac{\log(\frac{2}{a})}{2n}}) \leq 2\frac{a}{2} = a$

$\Leftrightarrow P(|\hat{p}_n - p| \geq \varepsilon_n) \leq a \Leftrightarrow P(|\hat{p}_n - p| \leq \varepsilon_n) \geq 1-a$

b) The plot shows the coverage probability as a function of the sample size on a logarithmic scale. The red dashed line represents the target coverage probability $1-a = 0.95$. As $n\uparrow \Rightarrow$ the coverage prob. approaches the desired level of 0.95. For smaller sample sizes (eg. n=10) the coverage prob is lower than 0.95 indicating that the confidence interval may not be as reliable for small n. As the sample size increases (n=1000 & n=10000) the coverage prob converges towards 0.95, demonstrating that the confidence interval becomes more accurate with larger sample sizes.

c) The plot shows the length of confidence intervals as a function of the sample size n with both axes on a logarithmic scale. Key observations:

- The length of the confidence interval decreases as n increases.
- This indicates that with larger sample sizes the confidence interval becomes narrower ~~with both~~ leading to more precise estimates of p.

d) The plot illustrates the probability of making a correct decision as a function of the sample size n when the true proportion has changed from $p=0.4$ to $p=0.5$. Key observations: For smaller sample sizes (eg. $n=10$ the probability of making a correct decision (i.e. deciding that $p=0.4$) is relatively high as the confidence interval is wider and more likely to contain $p=0.4$ by chance. As the sample size increases (eg $n=1000$ and beyond) the probability of making a correct decision decreases significantly. This is because larger sample sizes produce more accurate estimates around the true $p=0.5$ resulting in narrower confidence intervals that are less likely to contain $p=0.4$.

Chapter 6: Solutions

1) Let $X_1 \ldots X_n \sim \text{Poisson}(\lambda)$ and let $\hat{\lambda} = n^{-1}\sum_{i=1}^{n} X_i$. Find the bias, se, and MSE of this estimator.
Since $E_\lambda[\hat{\lambda}] = E_\lambda[X_1] \Rightarrow$ estimator unbiased. $se(\hat{\lambda})^2 = \frac{V_\lambda(X_1)}{n} = \frac{\lambda}{n}$. By the bias-variance decomposition the MSE is equal to $se(\hat{\lambda})^2$

2) Let $X_1 \ldots X_n \sim \text{Unif}(0, \vartheta)$ and let $\hat{\vartheta} = \max\{X_1, \ldots, X_n\}$. Find the bias, se and MSE of this estimator.
If y is between 0 and ϑ, $P_\vartheta(\hat{\vartheta} \le y) = P_\vartheta(X_1 \le y)^n = (y/\vartheta)^n$. ~~Hence~~ differentiating yields the PDF of $\hat{\vartheta}$ between 0 and ϑ as $y \to n(y/\vartheta)^n/y$. Therefore, $E_\vartheta[\hat{\vartheta}] = \int_0^\vartheta n(y/\vartheta)^n \, dy = \vartheta \cdot n/(n+1)$. It follows that the bias of this estimator is $-\vartheta/(n+1)$. Moreover, $se(\hat{\vartheta})^2 = \int_0^\vartheta ny(y/\vartheta)^n dy - E[\hat{\vartheta}]^2 = \vartheta^2 n/(n+2) - E[\hat{\vartheta}]^2$. By the bias-variance decomposition the MSE is: $\vartheta^2 n/(n+2) - \vartheta^2(n^2-1)/(n+1)^2$ $\quad \hat{\vartheta}(n+1)/n$ is an unbiased estimator

3) Let $X_1 \ldots X_n \sim \text{Uniform}(0, \vartheta)$ and let $\hat{\vartheta} = 2\bar{X}_n$. Find bias, se, MSE of the estimator.
Since $E_\vartheta[\hat{\vartheta}] = 2E_\vartheta[X_1] = \vartheta$: estimator is unbiased.
$se(\hat{\vartheta})^2 = 4V_\vartheta(X_1)/n = \vartheta^2/3n$. By the bias-variance decomposition the MSE is equal to $se(\hat{\vartheta})^2$

Chapter 5. Exercises:

4. Convergence in Probability: A sequence X_n converges in prob. to X (say $X=0$) for every $\varepsilon>0$: $\mathbb{P}(|X_n - 0| > \varepsilon) \to 0$ as $n\to\infty$ (prob X_n deviates from zero more ...)

case 1: $\varepsilon \geq 1$: $P(|X_n - 0| > \varepsilon) = P(X_n = n) = \frac{1}{n^2}$ as $n\to\infty$ $\frac{1}{n^2}\to 0$

case 2: $\varepsilon < 1$: $P(|X_n - 0| > \varepsilon) = P(X_n = n) = \frac{1}{n^2}$ as $n\to\infty$ $\frac{1}{n^2}\to 0$ $X_n \xrightarrow{P} 0$

Convergence in Quadratic Mean: A sequence X_n converges in quadratic mean to X (say $X=0$) if $E[(X_n - 0)^2] = E[X_n^2] \to 0$ as $n\to\infty$

$E[X_n^2] = \left(\frac{1}{n}\right)^2 P(X_n = \frac{1}{n}) + n^2 P(X_n = n)$

$= \left(\frac{1}{n}\right)^2\left(1-\frac{1}{n^2}\right) + n^2\cdot\frac{1}{n^2} = \frac{1}{n^2} - \frac{1}{n^4} + 1$ as $n\to\infty$ $E[X_n^2]\to 1$

Since $E[X_n^2]$ approaches 1 so X_n does not converge in quadratic mean

8) Poisson distribution with mean $\mu_Y = n\cdot\lambda = 100\cdot 1 = 100$

Instead of using Poisson we apply CLT to approximate Y by a Normal distribution:

$Y \sim N(\mu_Y, \sigma_Y^2)$ where $\mu_Y = \sigma_Y^2 = n\cdot\lambda = 100\cdot 1 \Rightarrow \sigma_Y = \sqrt{100} = 10$

$Z = \frac{Y-\mu_Y}{\sigma_Y} = \frac{90-100}{10} = -1$

$P(Y<90) = P(Z<-1) \approx 0.1587$

12) Convergence in Distribution: A sequence of RV X_n converges in distribution to a RV X $(X_n \xrightarrow{d} X)$ if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all $x\in\mathbb{R}$ where $F_X(x)$ continuous

For positive integer-valued RV the CDF $F_{X_n}(x)$ is defined as:

$F_{X_n}(x) = P(X_n \leq x) = \sum_{k\leq x} P(X_n = k)$ for integer k the continuity of CDF simplifies to checking the PMF. $P(X_n = k) \to P(X=k)$ for all $k\in\mathbb{Z}^+$

1. If $X_n \xrightarrow{d} X$ then $\lim_{n\to\infty} P(X_n = k) = P(X=k)$ for every k. It jumps only at integers k:

(size of jump): $P(X=k) = F_X(k) - F_X(k-1)$
similarly $P(X_n = k) = F_{X_n}(k) - F_{X_n}(k-1)$
} convergence $F_{X_n}(x) \to F_X(x)$

2. If $\lim_{n\to\infty} P(X_n = k) = P(X=k)$ for every k then $X_n \xrightarrow{d} X$:

$F_{X_n}(x) = \sum_{k\leq x} P(X_n = k)$, $F_X(x) = \sum_{k\leq x} P(X=k)$.

If $\lim_{n\to\infty} P(X_n = k) = P(X=k)$ for all k then

$\lim_{n\to\infty} \sum_{k\leq x} P(X_n = k) = \sum_{k\leq x} \lim_{n\to\infty} P(X_n = k) = \sum_{k\leq x} P(X=k)$

This implies $\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \Rightarrow X_n \xrightarrow{d} X$

0. Exercises

$X_1, X_2 \dots X_n \sim \text{Poisson}(\lambda)$ $\quad \hat{\lambda} = \sum n^{-1} \cdot \sum_{i=1}^{n} X_i$

$\text{Bias}(\hat{\lambda}) = \mathbb{E}[\hat{\lambda}] - \lambda = \mathbb{E}\left[\frac{1}{n} \sum X_i\right]$ (1)

since X_i: iid Poisson with mean $= \lambda$: $\mathbb{E}[X_i] = \lambda$ so $E[\sum X_i] = n \cdot \lambda$ (2)

adding into (1) → (2)

$$\mathbb{E}[\hat{\lambda}] = \frac{1}{n} \cdot n\lambda = \lambda \Rightarrow \text{Bias}(\hat{\lambda}) = \lambda - \lambda = 0$$

↳ $\hat{\lambda}$ unbiased estimator of λ

$\text{Var}(\hat{\lambda}) = \text{Var}\left(\frac{1}{n} \sum X_i\right) \overset{(1)}{\underset{\text{since } X \text{ independent}}{=}} \sum^{n} \text{Var}(X_i)$

$X \sim$ Poisson $\Rightarrow \text{Var}(X_i) = \lambda$ so $\text{Var}(\sum X_i) = n \cdot \lambda$ (2)

adding into (1) → (2) $\quad \text{Var}\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n^2} \text{Var}(\sum X_i) = \frac{1}{n^2} n\lambda = \frac{\lambda}{n}$

$\Rightarrow SE = \sqrt{\frac{\lambda}{n}}$

$\text{MSE}(\hat{\lambda}) = \text{Var}(\hat{\lambda}) + \text{Bias}(\hat{\lambda})^2 = \frac{\lambda}{n}$

(*) Property of Variance
$\text{Var}(cX) = c^2 \text{Var}(X)$

2. $X_1, X_2 \dots X_n \sim \text{Uniform}(0, \theta)$ $\quad \hat{\theta} = \max(X_1, \dots X_n)$

PDF ⇒ $f_X(x) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$ $\quad$ CDF $= F_X(x) \begin{cases} \frac{x}{\theta} & 0 \le x \le \theta \\ 0 & x < 0 \\ 1 & x > \theta \end{cases}$

Distribution of $\max\{X_1, \dots X_n\} = \hat{\theta}$

CDF $F_{\hat{\theta}}(x) = P(\hat{\theta} \le x) = P(X_1 \le x, X_2 \le x \dots X_n \le x)$

By independence $F_{\hat{\theta}}(x) = \prod^{n} P(X_i \le x) = (F_X(x))^n = \left(\frac{x}{\theta}\right)^n$, $0 \le x \le \theta$

PDF: differentiate CDF

$f_{\hat{\theta}}(x) = \frac{d}{dx} F_{\hat{\theta}}(x) = \frac{d}{dx}\left(\frac{x}{\theta}\right)^n = n \frac{x^{n-1}}{\theta^n}$, $0 \le x \le \theta$

$E[\hat{\theta}] = \int_0^{\theta} x f_{\hat{\theta}}(x)\, dx = \int_0^{\theta} x \cdot n \frac{x^{n-1}}{\theta^n} dx = n \int_0^{\theta} \frac{x^n}{\theta^n} dx = \frac{n}{\theta^n} \int_0^{\theta} x^n dx = \frac{n}{\theta^n} \cdot \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \theta$

Bias of θ: $= E[\hat{\theta}] - \theta = \frac{n}{n+1}\theta - \theta = -\frac{\theta}{n+1}$

$\text{Var}(\hat{\theta}) = E[\hat{\theta}^2] - (E[\hat{\theta}])^2$

$E[\hat{\theta}^2] = \int_0^{\theta} x^2 f_{\hat{\theta}}(x)\, dx = n \int_0^{\theta} \frac{x^{n+1}}{\theta^n} dx = \frac{n}{n+2} \theta^2$

↳ $\text{Var}(\hat{\theta}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1} \cdot \theta\right)^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 \Leftrightarrow SE(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$

$\text{MSE} = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2 = \text{Var}(\hat{\theta}) + \frac{\theta^2}{(n+1)^2}$

3) $X_1 \dots X_n \sim \text{Uniform}(0,\vartheta)$ $\quad \hat\vartheta = 2\bar X_n$

$\hat\vartheta = 2\bar X_n = 2\left(\frac{1}{n}\sum X_i\right)$

$E[X_i] = \frac{\vartheta}{2} = E[\bar X_n] \Leftrightarrow E[2\bar X_n] = \vartheta$ $\quad \text{Bias}(\hat\vartheta) = E(\hat\vartheta) - \vartheta = \vartheta - \vartheta = 0$ (un

$Var(X_i) = \frac{\vartheta^2}{12}$ for sample mean $Var(\bar X_n) = \frac{Var(X_i)}{n} = \frac{\vartheta^2}{12n} \Rightarrow Var(2\bar X_n) = \frac{\vartheta^2}{3n}$

$se(\bar X_n) = \sqrt{\frac{\vartheta^2}{12n}} \Rightarrow se(2\bar X_n) = \frac{\vartheta}{\sqrt{3n}}$

$MSE = Var(\hat\vartheta) = \frac{\vartheta^2}{3n}$

Chapter 9: Exercises:

1) Let $X_1, \dots X_n \sim \text{Gamma}(a,b)$ Find the method of moments estimator for a and b

$\mu = E[X] = \frac{a}{b}$ $\quad Var[X] = \frac{a}{b^2}$

$\hat\mu = \frac{1}{n}\sum X_i$ $\quad \hat\sigma^2 = \frac{1}{n-1}\sum(X_i - \hat\mu)^2$ (equating the sample moments to the population moments)

$\mu = \frac{a}{b} \Rightarrow \hat\mu = \frac{a}{b} \Rightarrow a = b\cdot\hat\mu$ (1) $\quad Var(X) = \frac{a}{b^2} \Rightarrow \hat\sigma^2 = \frac{a}{b^2}$ (2)

(1)(2): $\frac{b\hat\mu}{b^2} = \hat\sigma^2 \Leftrightarrow \frac{\hat\mu}{b} = \hat\sigma^2 \Leftrightarrow \hat b = \frac{\hat\mu}{\hat\sigma^2}$

$a = b\hat\mu = \frac{\hat\mu}{\hat\sigma^2}\cdot\hat\mu = \frac{\hat\mu^2}{\hat\sigma^2} \Leftrightarrow \quad \hat a = \frac{\hat\mu^2}{\hat\sigma^2}$

2) Let $X_1 \dots X_n \sim \text{Uniform}(a,b)$ where a and b are unknown parameters and $a<b$.

a) Find the method of moments estimators for a and b

$E[X] = \frac{a+b}{2}$ $\quad Var(X) = \frac{(b-a)^2}{12}$ $\quad \hat\mu = \frac{a+b}{2} \Leftrightarrow a+b = 2\hat\mu$ ①

$\frac{(b-a)^2}{12} = \hat\sigma^2 \Leftrightarrow b-a = \sqrt{12\hat\sigma^2}$ ②

①+②: $(a+b)+(b-a) = 2\hat\mu + \sqrt{12}\,\hat\sigma \Leftrightarrow \hat b = \hat\mu + \frac{\sqrt{12}}{2}\hat\sigma$

$\hat a = \hat\mu - \frac{\sqrt{12}}{2}\hat\sigma$

b) Find MLE $\hat a$ and $\hat b$

$f(x,a,b) = \frac{1}{b-a}$ $\quad a \le x \le b$

$L(a,b) = \prod^n f(x_i, a, b) = \prod^n \frac{1}{b-a} = \frac{1}{(b-a)^n}$ $\quad a \le \min(X_1, X_2, \dots X_n)$ 1) $\quad b \ge \max(X_1, X_2, \dots X_n)$ 2)

$\log L(a,b) = -n\log(b-a)$

To maximise the log likelihood function we need to consider that 1) & 2)

- Maximising wrt b: To max $\log L(a,b) = -n\log(b-a)$ we want to minimise $b-a$. Since $b \ge \max(X_1, X_2 \dots X_n)$ the smallest value of b that satisfies this constraint is $b = \max(X_1, X_2 \dots$

- Maximising wrt a: To max log we want to minimise $b-a$, so we choose $a = \min(X_1 \dots X_n)$ because $a \le \min(X_1 \dots X_n)$

$\hat a = \min(X_1, X_2, \dots X_n)$ & $\hat b = \min(X_1, X_2 \dots X_n)$ MLE for a and b

c) Let $\tau = \int x\, dF(x)$ Find the MLE of τ

$\tau = \int x\, dF(x) = E[X]$

$L(\vartheta \mid x_n) = \prod_{i=1}^{n} f(x_i;\vartheta)$

$\ell(\vartheta) = \sum \log f(x_i,\vartheta)$

...imate T, we need first to find MLE of ϑ by max. the log likelihood functi...

$\hat{\vartheta} = \arg\max_{\vartheta} \ell(\vartheta)$

...nce we have $\hat{\vartheta}$ the MLE of T is given by the distribution $F(x,\hat{\vartheta})$ that depends on the ...ent of distribution if $F(x)$: non-parametric the expected value is simply the sample mean:

$\hat{T} = \frac{1}{n}\sum_{i=1}^{n} X_i$, this is the MLE of T when no parametric assumption about $F(x)$ is ma...

...'s answer:

$T = E[X] = \frac{a+b}{2}$ MLEs are ~~equivalent~~ equivariant under transformations if $T = g(a,b) = \frac{a+b}{2}$ then the MLE of T is obtained by substituting MLEs of a and b into $g(a,b)$. Thus $\hat{T} = g(\hat{a},\hat{b}) = \frac{\hat{a}+\hat{b}}{2}$

$\hat{T} = \frac{\hat{a}+\hat{b}}{2} = \frac{\min(X_1 \dots X_n) + \max(X_1 \dots X_n)}{2}$

d) Let $\hat{T}$ be the MLE of T. Let $\tilde{T}$ be a nonparametric plug-in estimator of $T = \int x\,dF(x)$. Suppose that $a=1$, $b=3$, $n=10$. Find the MSE of $\hat{T}$ by simulation. Find the MSE of $\tilde{T}$ analytically. Compare.

MSE of $\tilde{T}$ (nonparametric plug-in estimator is the sample mean $\tilde{T}$ that is unbiased

$Var(\tilde{T}) = \frac{Var(X)}{n}$ $\quad Var(X) = \frac{(b-a)^2}{12} = \frac{(3-1)^2}{12} = \frac{1}{3}$

$\hookrightarrow Var(\tilde{T}) = \frac{1}{3n} = \frac{1}{30}$

since $\tilde{T}$ unbiased $MSE(\tilde{T}) = \frac{1}{30} = Var(\tilde{T}) = 0.033$

By the simulation the $MSE(\hat{T}) = 0.0151$

3) Let $X_1 \dots X_n \sim N(\mu,\sigma^2)$. Let T be the 0.95 percentile i.e. $P(X<T) = 0.95$

a) Find the MLE of T.

$P(X<T) = 0.95$ using CDF of the normal distribution we have: $T = \mu + z_{0.95}\cdot\sigma$

$L(\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$

$\ell(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad \Rightarrow \hat{T} = \bar{X} + z_{0.95}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2} = \hat{\mu} + \overset{1.645}{z_{0.95}}\,\hat{\sigma}$

b) Find an expression for an approximate $1-\alpha$ CI for T

variance of $\hat{T}$: $Var(\hat{T}) = Var(\hat{\mu}) + z^2_{0.95}\,Var(\hat{\sigma}) \qquad Var(\hat{\mu}) = \frac{\sigma^2}{n} \quad Var(\hat{\sigma}) \approx \frac{\sigma^2}{2n}$

$SE(\hat{T}) = \sqrt{Var(\hat{\mu}) + z^2_{0.95}\,Var(\hat{\sigma})}$

$CI = \left(\hat{T} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\sigma}^2}{n} + z^2_{0.95}\frac{\hat{\sigma}^2}{2n}}\right)$ using Delta method

c) Suppose the data are: [3.23, -2.50, 1.88, -0.68, 4.43, 0.17, 1.03, -0.07, -0.01, 0.76, 1.76, 3.18, 0.33, -0.31, 0.30, -0.61, 1.52, 5.43, 1.54, 2.28, 0.42, 2.33, -1.03, 4, 0.39]

Find the MSE $\hat{T}$. Find the SE using the delta method. Find the SE using the parametric bootstrap (computer).

Chapter 23 - Exercises: Probability Redux: Stochastic Processes

1) Let $X_0, X_1, \ldots$ be a MC with states $\{0,1,2\}$ and transition matrix:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.9 & 0.1 & 0.0 \\ 0.1 & 0.8 & 0.1 \end{bmatrix} \end{array}$$

Assume that $\mu_0 = (0.3, 0.4, 0.3)$ Find $P(X_0 = 0, X_1 \ldots$ and $P(X_0 = 0, X_1 = 1, X_2 = 1)$. initial distribution

$$P(X_0 = 0, X_1 = 1, X_2 = 2) = \mu_0(0) \cdot P_{0,1} \cdot P_{1,2}$$
$$= 0.3 \cdot 0.2 \cdot 0 = 0$$

$$P(X_0 = 0, X_1 = 1, X_2 = 1) = \mu_0(0) \cdot P_{0,1} \cdot P_{1,1}$$
$$= 0.3 \cdot 0.2 \cdot 0.1 = 0.006$$

2) Let $Y_1, Y_2, \ldots$ be a sequence of iid observations s.t. $P(Y=0) = 0.1$, $P(Y=1) = 0.3$, $P(Y=2) = 0.2$ and $P(Y=3) = 0.4$. Let $X_0 = 0$ and let $X_n = \max\{Y_1, \ldots, Y_n\}$. Show that $X_0, X_1, \ldots$ is a MC and find the transition matrix.

First we need to verify the Markov property:

$P(X_{n+1} = x \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+1} = x \mid X_n = x_n)$ → future state depends only on the current state and not on the entire history.

Given $X_n = \max\{Y_1, Y_2, \ldots, Y_n\}$ the next value X_{n+1} is determined by the maximum of

$X_{n+1} = \max(X_n, Y_{n+1})$ This shows that the value of X_{n+1} depends only on X_n and the new observation Y_{n+1}, meaning the future state depends only on the present state.

If $X_n = i$ the value of $X_{n+1} = \max\{i, Y_{n+1}\}$ (depends on Y_{n+1}).

If $i = 0$:
$X_{n+1} = 0$ if $Y_{n+1} = 0$
$X_{n+1} = 1$ if $Y_{n+1} \in \{1,2,3\}$
$X_{n+1} = 2$ if $Y_{n+1} \in \{2,3\}$
$X_{n+1} = 3$ if $Y_{n+1} \in \{3\}$

Thus:
$P(X_{n+1} = 0 \mid X_n = 0) = P(Y_{n+1} = 0) = 0.1$
$P(X_{n+1} = 1 \mid X_n = 0) = P(Y_{n+1} \in \{1,2,3\}) = 0.3 + 0.2 + 0.4 = 0.9$
$P(X_{n+1} = 2 \mid X_n = 0) = P(Y_{n+1} \in \{2,3\}) = 0.2 + 0.4 = 0.6$
$P(X_{n+1} = 3 \mid X_n = 0) = P(Y_{n+1} \in \{3\}) = 0.4$

If $i = 1$:
$X_{n+1} = 1$ if $Y_{n+1} \in \{0,1\}$
$X_{n+1} = 2$ if $Y_{n+1} \in \{2,3\}$
$X_{n+1} = 3$ if $Y_{n+1} \in \{3\}$

Thus:
$P(X_{n+1} = 1 \mid X_n = 1) = P(Y_{n+1} \in \{0,1\}) = 0.1 + 0.3 = 0.4$
$P(X_{n+1} = 2 \mid X_n = 1) = P(Y_{n+1} \in \{2,3\}) = 0.2 + 0.4 = 0.6$
$P(X_{n+1} = 3 \mid X_n = 1) = P(Y_{n+1} = 3) = 0.4$

If $i = 2$:
$X_{n+1} = 2$ if $Y_{n+1} \in \{0,1,2\}$
$X_{n+1} = 3$ if $Y_{n+1} \in \{3\}$

Thus:
$P(X_{n+1} = 2 \mid X_n = 2) = P(Y_{n+1} \in \{0,1,2\}) = 0.1 + 0.3 + 0.2 = 0.6$
$P(X_{n+1} = 3 \mid X_n = 2) = P(Y_{n+1} = 3) = 0.4$

If $i = 3$: $X_{n+1} = 3$ Y_{n+1} is any value the max value will always be 3.
$P(X_{n+1} = 3 \mid X_n = 3) = 1$

$$P = \begin{bmatrix} 0.1 & 0.4 & 0.6 & 0.4 \\ 0.4 & 0.4 & 0.6 & 0.4 \\ 0.6 & 0.4 & 0.6 & 0.4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

... two-state MC with states $\mathcal{X} = \{1, 2\}$ and transition matrix:

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$ where $0 < a < 1$ and $0 < b < 1$. Prove that:

$$\lim_{n \to \infty} P^n = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$$

MC reaches ~~the~~ steady state (if chain is irreducible & aperiodic), the stationary distribution $\pi = [\pi_1, \pi_2]$ satisfies $\pi P = \pi$.

$$[\pi_1 \ \pi_2] \cdot \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} \quad \left\{ \begin{array}{l} \pi_1(1-a) + \pi_2 \cdot b = \pi_1 \\ \pi_1 \cdot a + \pi_2(1-b) = \pi_2 \\ \pi_1 + \pi_2 = 1 \end{array} \right.$$

after solving the system we get

$$\pi_1 = \frac{b}{a} \cdot \pi_2 \quad \Rightarrow \quad \pi = \left[\frac{b}{a+b} \quad \frac{a}{a+b} \right]$$

* Key property of MC is that as $n \to \infty$ the powers of the transition matrix P^n converges to a matrix where each row is the stationary distribution (if irreducible & aperiodic): $\lim_{n \to \infty} P^n = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$

(4) Consider the chain from question 3 and set $a = 0.1$ and $b = 0.3$. Simulate the chain. Let: $\hat{p}_n(1) = \frac{1}{n} \sum^n I(X_i = 1)$, $\hat{p}_n(2) = \frac{1}{n} \sum^n I(X_i = 2)$ be the proportion of times the chain is in state 1 and state 2. Plot $\hat{p}_n(1)$ and $\hat{p}_n(2)$ versus n and verify that they converge to the values predicted from the answer in the previous question. (computer-based)

(5) An important MC is the branching process which is used in biology, genetics, nuclear physics and many other. Suppose that an animal has Y children. ~~..~~ Let $p_k = P(Y = k)$. Hence $p_k \geq 0$ for all k and $\sum_{k=0}^{\infty} p_k = 1$. Assume each animal has the same lifespan and that they produce offspring according to the distribution p_k. Let X_n be the # of animals in the n^{th} generation. Let $Y_1^{(n)}$, ~~...~~ $Y_{X_n}^{(n)}$ be the offspring produced in the n^{th} generation. Note that $X_{n+1} = Y_1^{(n)} + \dots + Y_{X_n}^{(n)}$.
Let $\mu = E(Y)$ and $\sigma^2 = V(Y)$. Assume that $X_0 = 1$. Let $M(n) = E(X_n)$ and $V(n) = V(X_n)$.
a) Show that $M(n+1) = \mu \cdot M(n)$ and $V(n+1) = \sigma^2 M(n) + \mu^2 \cdot V(n)$ [computer]

(6) Let $P = \begin{bmatrix} 0.40 & 0.50 & 0.10 \\ 0.05 & 0.70 & 0.25 \\ 0.05 & 0.50 & 0.45 \end{bmatrix}$ Find stationary distribution π.

$$\left. \begin{array}{l} 0.4\pi_1 + 0.05\pi_2 + 0.05\pi_3 = \pi_1 \\ 0.50\pi_1 + 0.70\pi_2 + 0.50\pi_3 = \pi_2 \\ 0.10\pi_1 + 0.25\pi_2 + 0.45\pi_3 = \pi_3 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{array} \right\} \quad \pi_1 = 0.0769, \ \pi_2 = 0.625, \ \pi_3 = 0.2981$$

(7) Show that if i is a recurrent state $i \leftrightarrow j$, then j is a recurrent state [computer]

(8) Let

$$P = \begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

which states are transient? Which are recurrent?

State 1: can transition to 1, 3, 6 (transient)
State 2: ~~possible transitions~~ 1, 2, 3 (transient)
State 3: absorbing state has 1 in the position (3,5) $\Rightarrow$ (recurrent)
State 4: possible transitions: 1, 2, 3, 6 (transient)
State 5: only transition is 3 that is absorbing (transient)
State 6: no outgoing transitions to other states (recurrent & absorbing)

Let $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ show that $\pi = (1/2, 1/2)$ is a stationary distribution. Does this chain converge? Why, why not?

$$(\pi_1 \quad \pi_2)\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} \qquad \left.\begin{array}{l} \pi_1 = \pi_2 \\ \pi_2 = \pi_1 \\ \pi_1 + \pi_2 = 1 \end{array}\right\} \pi = \left(\tfrac{1}{2} \ \tfrac{1}{2}\right).$$

To see it it converges we need to examine the behaviour of the system as it evolves over time. Examine the powers of P:

$P.P = P^2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow$ Identity matrix. $\Rightarrow$ meaning after 2 steps the system returns to the original state. This shows that the MC oscillates between the two states never settling into a single state. The system alternates between state 1 and 2, with each time step.

The MC does not converge to a single state, but the stationary distribution is stable in the long run.