All of Statistics

sagisk

http://www.stat.cmu.edu/ larry/=stat325.01/sol11.pdf

Chapter 1: Probability

1.1

Suppose $w \in \cup_{i=1}^{n} A_i$ for some n, then by the construction we know that $w \in B_i, i \in [1, n]$, so $w \in \cup_{i=1}^{n} B_i$ and $\cup_{i=1}^{n} A_i \subset \cup_{i=1}^{n} B_i$. If $w \in \cup_{i=1}^{n} B_i$ then $w \in A_j \setminus \cup_{k=1}^{j-1} A_k$. Therefore, $w \in \cup_{i=1}^{n} A_i$ and $\cup_{i=1}^{n} B_i \subset \cup_{i=1}^{n} A_i$.

1.2

$\emptyset \cup \Omega = \Omega$, so they are disjoint, $P(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega)$ by the Axiom 3, but then by the Axiom 2 we know that $P() = 1$, then $P(\emptyset \cup \Omega) = P(\Omega) = 1 = P(\emptyset) + P(\Omega)$, so $P(\emptyset) = 0$.

$A \subset B, B = A \cap B \cup B \setminus A = A \cup B \setminus A$, the constructed parts are disjoint so, by the Axiom 3, $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

$0 \leq P(A) \leq 1$. $P(A) \geq 0$ by the Axiom 1. Also, by the definition $A \subset \Omega$, therefore by the previous point $P(A) \leq P(\omega) = 1$.

$A^c = \{w \in \Omega : w \notin A\}$ by definition. So, $A^c \cap A = \{\emptyset\}$. Also it can be shown that $A^c \cup A = \Omega$. By the Axiom 2 and Axiom 3, $P(A \cup A^c) = P(\Omega) = 1 = P(A) + P(A^c), P(A^c) = 1 - P(A)$.

$A \cap B = \emptyset$, so by the Axiom 3 the given relation is true.

1.3

a) $B_n = \cup_{i=n}^{\infty} A_i$. Lets take some $k, j \in N$ such that $k < j$. Then $B_k = \cup_{i=k}^{\infty} A_i$ and $B_j = \cup_{i=j}^{\infty} A_i$. If $w \in B_j$, then $w \in \cup_{i=j}^{\infty} A_i$. Since $\cup_{i=j}^{\infty} A_i \in B_k$, then $w \in B_k$, so $B_j \subset B_k$. The converse is not true, since if $w \in A_k$ it is in B_k but it is not in B_j.

$C_n = \cap_{i=n}^{\infty} A_i$. Lets take some $k, j \in N$ such that $k < j$. $C_k = \cap_{i=k}^{\infty} A_i$ and $C_j = \cap_{i=j}^{\infty} A_i$. If $w \in C_k$, it must must be in C_j, since $w \in A_i$ for all $i \in [k, \infty)$. So, $C_k \subset C_j$. If $w \in C_j$ then it must not be in C_n, since $w \notin A_k$.

b) If $w \in \cap_{n=1}^{\infty} B_n$, then $w \in \cap_{n=1}^{\infty} \cup_{i=n}^{\infty} A_i$. Since for $k < j$, $B_j \subset B_k$, then $B_k \cap B_j = B_j$. (https://math.stackexchange.com/questions/1166815/infinite-number-of-events-and-an-element-of-their-intersection)

c)

1.4

Let's prove by induction on n. The base case: $n = 1$, then $A_1^c = A_1^c$ trivially holds. Inductive step: Suppose the statement is true for $I_0 = \{1, 2, ...n - 1\}$, now let $I_0 =$

$\{1, 2, ...n-1, n$ $(\cup_{i\in I} A_i)^c = (\cup_{i\in I_0} A_i \cup A_n)^c = (B \cup A_n)^c = B^c \cap A_n^c$ by the inductive assumption. Now, $B^c \cap A_n^c = \cap_{i\in I_0} A^c \cap A_n^c = \cap_{i\in I} A_i^c$.

Now for arbitrary n. (I understand it as $n \to \infty$. Am I right?) If $x \in (\cup_1^\infty A_i)^c$, then $x \notin \cup_{i\in I} A_i$. This means, that for all $n \geq 1$ $x \notin A_n$, which means that it is in $(A_n)^c$ for every $n \geq 1$. Therefore $x \in \cap_1^\infty A_n^c$.

If $x \in \cap_1^\infty A_n^c$, then $x \in A_n^c$ for all $n \geq 1$, so $x \notin A_n$ for any $n \geq 1$, which means that $x \notin \cup_1^\infty A_n$, so $x \in (\cup_1^\infty A_n)^c$

1.5

The sample space $S = \{w_1, w_2, w_3, ...\}$ where $w_i = \{H, T\}$ for $j = i+1, w_k = w_i = H$. If k tosses are required that means that in the first $k-1$ tosses we managed to get exactly one H while the others are T and the k-th toss is a H.

$$(k-1)!/(1! * (k-2)!)P(H)P(T)^{k-2} * P(H)$$

1.6

Suppose this can be done i.e. $P(A) = P(B)$ for any $|A| = |B|$. Let $A = [0, 1), B = [1, 2)...$, Since the events are disjoint then by the Axiom 3 $P(A \cup B \cup C...) = \sum_{i=1}^{\infty} P(A) = P(A)\sum_{i=1}^{\infty}$ where the sum diverges. However, we showed that for any event H, $P(H) \leq 1$. So, a contradiction.

1.7

Let $B_n = A_n \setminus \cup_{i=1}^{n-1} A_i$. Let $k < j$. Then let $w \in B_k \cap B_j$, then $w \in A_k \setminus \cup_{i=1}^{k-1} A_i$ and $w \in A_j \setminus \cup_{i=1}^{j-1} A_i$. However $A_k \subset \cup_{i=1}^{j-1} A_i$, so, it is not possible, they are disjoint. By the similar argument to the one given in exercise 1, it can be shown that $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} B_i$. Then by the Theorem of continuoty of probabilities, Axiom 3 and the fact that the probability of the subset of the set is smaller or equal than the probability of the set. $P(A) = P(\cup_{i=1}^{\infty} A_i) = P(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) = P(B_1) + P(B_2) + ... \leq P(A_1) + P(A_2) + ...$

1.8

By the De Morgan's law (exercise 4) $(\cap_{i=1}^{\infty} A_i)^c = \cup_{i=1}^{\infty} A_i^c$. By the exercise 2 we know that for any event $B, P(B) = 1 - P(B^c)$. So, $P(\cap_{i=1}^{\infty} A_i) = 1 - P((\cap_{i=1}^{\infty} A_i)^c) = 1 - P(\cup_{i=1}^{\infty} A_i^c)$, where $P(A_i^c) = 0$ for all $i \in [1, \infty)$. Moreover, by the previous exercise $P(\cup_{i=1}^{\infty} A_i^c) \leq \sum_{i=1}^{\infty} P(A_i) = 0$, this in combination with the exercise 2 gives that $P(\cup_{i=1}^{\infty} A_i^c) = 0$ and $P(\cap_{i=1}^{\infty} A_i) = 1$.

1.9

$P(*|B) = P(* \cap B)/P(B)$, since we have that for any event $A, P(A) \geq 0$ by the Axiom 1, then $P(*|B) \geq 0$. $P(\Omega|B) = P(\Omega \cap B)/P(B) = P(B)/P(B) = 1$, since B is a subset of

Ω, so Axiom 2 holds. Let A_i be disjoint events. $P(\cup_{i=1}^{\infty} A_i|B) = P(\cup_{i=1}^{\infty} A_i \cap B)/P(B) = P(A_1 \cap B \cup A_2 \cap B \cup ...)/P(B) = \sum_{i=1}^{\infty} P(A_i \cap B)/P(B)$.

To Do: Show that $\cup_{i=1}^{\infty} A_i|B = \cup_{i=1}^{\infty} A_i \cap B$. Show that $A_i \cap B$'s are disjoint.

1.10

$P(M) = 1/2, P(P|M) = P(M|P)P(P)/P(M) = (1/2 * 1/3)/(1/2) = 1/3$, so the probability that the prize is under door 1 is 1/3. Therefore, by switchin it is equal to 2/3.

1.11

$P(AB) = P(A)P(B)$ $A^c \cap B^c = (A \cup B)^c$ $P((A \cup B)^c) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A)P(B) = 1 - P(A) - P(B)(1 - P(A)) = P(A^c) - P(B)P(A^c) = P(A^c)(1 - P(B)) = P(A^c)P(B^c)$

1.12

$P(Other_G|Saw_G) = P(Other_G \cap Saw_G)/P(Saw_G)$

$P(Other_G \cap Saw_G) = 1/3$ $P(Saw_G) = P(Saw_G|BothGreen)P(BothGreen) + P(Saw_G|BothRed)P(B$ $P(Saw_G|GreenAndRed)P(GreenAndRed) = 1 * 1/3 + 0 * 1 + 1/2 * 1/3 = 1/3 + 1/6 = 3/6 = 1/2$ $P(Other_G|Saw_G) = (1/3)/(1/2) = 2/3$

1.13

(h,h,t), (t,t,h) $P(3Tosses) = 2/8 = 1/4$

1.14

If $P(A) = 0$, then $P(A \cap B) = P(B|A)P(A) = 0 = P(A)P(B)$. If $P(A) = 1, P(A|B) = P(A \cap B)/P(B) = P(B|A)P(A)/(P(B|A)P(A) + P(B|A^c)P(A^c)) = P(B|A)/P(B|A) = 1 = P(A)$.

Suppose $P(A)$ is equal to some other number in range $(0, 1)$, then by the conditional probability $P(A) = P(A|A) = P(A)/P(A) = 1$, a contradiction.

1.15

a) $P(2or3blue|1or2or3blue) = P(2or3blueAND1or2or3blue)/P(1or2or3blue) = P(2or3blue)/P(1or2or$

$P(1or2or3blue) = P(1) + P(2) + P(3) - P(1and2) - P(1and3) - P(2and3) + P(1and2and3) = 3/4 - 3/16 + 1/64 = 37/64$ $P(2or3blue) = P(2blue) + P(3blue) - P(2and3blue) = C_2^3 * (1/16) * (3/4) + C_3^3 1/64 - 0 = 9/64 + 1/64 = 10/64$

$P(2or3blue|1or2or3blue) = (10/64)/(37/64) = 10/37$

b) $P(2or3blue|youngestblue) = P(2or3blueANDyoungestblue)/P(youngestblue)$

$P(2or3blueANDyoungestblue) = C_1^2 * 1/4 * 3/4 + C_2^2 * 1/16 = 6/16 + 1/16 = 7/16$ I think that $P(youngestblue) = 1$. However, I am not convinced why.

1.16

If A, B are independent: $P(A|B) = P(AB)/P(B) = P(A)P(B)/P(B) = P(A)$.

$P(A|B) = P(AB)/P(B)$ but also $P(B|A) = P(BA)/P(A) = P(AB)/P(A)$, (since $A \cap B = B \cap A$) so, the given relation holds.

1.17

$P(ABC) = P(A|BC)P(BC) = P(A|BC)P(B|C)P(C)$ by the repeated application of the Lemma 1.14.

1.18

$P(A_1|B) = P(A_1 \cap B)/P(B) = P(B|A_1)P(A_1)/P(B) < P(A_1)$ i.e. $P(B|A_1)P(A_1) < P(B)P(A_1)$. Suppose, $P(A_i|B) \leq P(A_i)$ for all $i = 2, ..k$. Then by the law of total probability (Theorem 1.16) $P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i) < \sum_{i=1}^{k} P(B)P(A_i) = P(B)\sum_{i=1}^{k} P(A_i) = P(B)$, so $P(B) < P(B)$, a contradiction.

Why $<$ rather then $\leq$? Because $P(A_1|B) < P(A_1)$ not less then equal and even if for all of the other i's they are equal the sign will remain $<$.

1.19

$P(W|V) = P(W \cap V)/P(V) =$

By the law of total probability (Theorem 1.16):

$P(V) = P(V|W)P(W) + P(V|M)P(M) + P(V|L)P(L) = 0.82*0.5 + 0.65*0.3 + 0.5*0.2$
$P(W \cap V) = P(V|W)P(W) = 0.82 * 0.5$.

1.20

a) $P(C_1|H) = 0, P(C_2|H) = P(H|C_2)P(C_2)/P(H) = 1/4 * 1/5/P(H)$

By the law of total probability (Theorem 1.16):

$P(H) = \sum_{i=1}^{5} P(H|C_i)P(C_i) = 1/5 * (0 + 1/4 + 1/2 + 3/4 + 1) = 1/5 * (5/2) = 1/2$

b) $P(H_2|H_1) = P(H_2 \cap H_1)/P(H_1)$

$P(H_2 \cap H_1) = \sum_{i=1}^{5} P(H_2 \cap H_1 \cap C_i) = P(H_2|H_1C_i)P(H_1|C_i)P(C_i)$ by the exercise 17.

Chapter 2: Random Variables

2.1

If the r.v. X is discrete then we have according to the theorem 2.8 that $F(x)$ is only right-continous. Which means that while approaching from the left there can be so called 'jumps'. Let x, y be the values of the r.v., such that $y < x$. Then $F(x^+) = F(x)$, while $F(x^-) = F(y-x)$. Moreover, $P(X = x) = \sum_{x_i \leq x} P(X = x) - \sum_{x_i \leq y} P(X = x) = P(X = x)$.

If the r.v. is continuous then the $F(x^+) = F(x^-)$. So, $P(X = x) = 0$.

2.2

2.3

1) Holds by the exercise 2.1.

2) $P(x < X \leq y) = \sum_x^y P(X = x) - P(X = x) = \sum_{x_i < y} P(X = x) - \sum_{x_i < x} P(X = x)$ The same way if the r.v. is contious.

3) $P(X > x) = \int_x^\infty f_X(x)dx$ Since $\int_{-\infty}^\infty f_X(x)dx = 1$, then $P(X > x) = \int_x^\infty f_X(x)dx = 1 - \int_{-\infty}^x f_X(x)dx$

2.4

2.5

By the definition we know that X, Y are independent if, for every $A, B, P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$. The if direction: Suppose that X, Y are independent but there exists some x, y such that $f_{XY}(x, y) \neq f_X(x)f_Y(y)$. However, then we can find subsets of the sample space namely the singleton subsets of $A = \{x\}, B = \{y\}$, such that $P(X \in A, Y \in B) \neq P(X \in A)P(Y \in B)$, which is a contradiction.

The only if direction: $f_X, Y(x, y) = f_X(x)f_Y(y)$ for all x, y from the sample space. Which means that in particular for all $A, B \subset \Omega$, if $x_0 \in A, y_0 \in A$, $f_X, Y(x_0, y_0) = f_X(x_0)f_Y(y_0)$, because x_0, y_0 are also in Ω. Therefore, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for all A, B. Or, $P(X \in A, Y \in B) = \sum_{x \in A} \sum_{y \in B} f_{X,Y}(x, y) = \sum_{x \in A} \sum_{y \in B} f_X(x)f_Y(y) = \sum_{x \in A} f_X(x) \sum_{y \in B} f_Y(y) = P(X \in A)P(Y \in B)$.

2.6

$P(Y=0) = P(I_A(x)=0) = P(X \in A^c) = p_{a_c}$ and $P(Y=1) = P(I_A(x)=1) = P(X \in A) = p_a$

So, $F_Y(y) = 0$ if $y < 0$, $F_Y(y) = p_{a_c}$ if $0 \leq y < 1$ and $F_Y(y) = 1$ if $y \geq 1$.

2.7

$X, Y\ ^{i.i.d.} Unif(0,1)$. $Z = min(X,Y)$. $P(Z > z) = P(min(X,Y) > z) = P(X > z \cap Y > z) = P(Y > z|X > z)P(X > z) = (1-z)^2$ by the independence and Bayes theorem. $P(Z \leq z) = 1-(1-z)^2, f_Z(z) = 2(1-z)$

2.8

$F_X(x) = P(X \leq x)$, $Y = max0, X$, $P(Y \leq y) = P(max0, X \leq y) = P(0 \leq y \cap X \leq y)$

Let $I_0(y) = 1$ if $y \geq 0$ and $I_0(y) = 0$ if $y < 0$. Then $F_Y(y) = P(Y \leq y) = 0$ if $y < 0$, $F_Y(y) = P(Y \in [0,1))$ and $F_Y(y) = P(1)$ if $y \geq 1$.

$P(0 \leq y \cap X \leq y) = P(X \leq y|y \geq 0)P(y \geq 0) = P(X \leq y)P(Y \geq 0) = F_X(y)(1 - P(Y \leq 0) = F_X(y)$

So, $F_Z(z) = 0$ if $y < 0$ and $F_Z(z) = F_X(y)$ if $y \geq 0$.

2.9

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

$F(x) = \int_0^x 1/\beta e^{-x/\beta} dx = 1/\beta \int_0^x e^{-x/\beta} dx$

Let $u = -x/\beta, du = -dx/\beta$, $1/\beta \int e^{-x/\beta} dx = -\int e^u du = -e^u = -e^{-x/\beta}$.

$F(x) = -e^{-x/\beta}|_0^x = (1 + e^{-x/\beta})$ Should be $1 - e^{-x/\beta}$. WHY???????

Let $F(q) = a = 1 - e^{-q/\beta}$, then $1 - a = e^{-q/\beta}, ln(1-a) = -q/\beta, q = -\beta ln(1-a)$.

2.10

If, g and h are invertable, $f_{g(X),h(Y)}(x,y) = P(g(X) = x|h(Y) = y)P(h(Y) = y) = P(X = g^{-1}x|y = h^{-1}y)P(Y = h^{-1}y) = P(X = x)P(Y = y)$, since by the definition of independence $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x, y.g(x) = y, h(y) = y

P(Y = m $\cap N = n)a) I one toss the sample space \Omega = \{H, T\}$. $X = 0$ if T, $X = 1$ if H and $Y = 0$ if H, $Y = 1$ if T. $P(X = 0 \cap Y = 0) = 0 \neq P(X = 0)P(Y = 0) = 1/4$.

b) $N\ Pois(\lambda), N = e^{-\lambda}\frac{\lambda^x}{x!}$ $P(X = k \cap N = n) = P(Y = n-k \cap N = n) = P(X = k|N = n)P(N = n) = (C_k^n 1/2^k * 1/2^{n-k}) * e^{-\lambda}\frac{\lambda^n}{n!} = \frac{e^{-\lambda}\lambda^n(1/2)^n}{k!}$

Let $k + m = n$ $P(X = k \cap Y = m \cap N = n) = P(X = k|Y = m \cap N = n)P(Y = m \cap N = n) = P(X = n-m|N = n)P(Y = m \cap N = n) = P(X = k|N = n)P(Y = m|N = n)P(N = n)$?????

2.12

$\int_0^\infty int_0^\infty f_{X,Y}(x,y)dydx = int_0^\infty int_0^\infty g(x)h(y)dydx = int_0^\infty g(x)dx int_0^\infty h(y)dy$

$f(x) = \int_0^\infty f(x,y)dy = g(x)\int_0^\infty h(y)dy$
$f(y) = \int_0^\infty f(x,y)dx = h(y)\int_0^\infty g(x)dx$
$\int_0^\infty f(x) = 1 = \int_0^\infty g(x)dx \int_0^\infty h(y)dy$ Putting all together we get that they are independent.

2.13

$P(Y \leq y) = P(e^X \leq y) = P(X \leq lny) = \Phi(lny),\ f_Y(y) = 1/y\Phi'(lny) = 1/y * f_X(lny) = (1/y) * 1/(2\pi)e^{-(lny^2)/2}$

2.14

$P(R \leq r) = P(\sqrt{X^2+Y^2} \leq r) = P(X^2 + Y^2\ ler^2) = \frac{areaofdiskofradiusr}{areaofdiskofradius1} = r^2.$

2.15

$Y = F(X), P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}y) = F(F^{-1}y) = y, Y\ Unif(0,1)$. Let $X = F^{-1}U, P(X \leq x) = P(F^{-1}U \leq x) = P(U \leq F(x)) = F_U(F(x)) = F(x)$ since U is uniform.

2.16

$P(X = k|X + Y = n) = P(X + Y = n|X = k)P(X = k)/P(X + Y = n)$
$P(X + Y = n|X = k)P(X = k) = P(Y = n - k|X = k)P(X = k) = e^{-u}\frac{u^{n-k}}{(n-k)!}e^{-\lambda}\frac{\lambda^k}{k!}$
$P(X + Y = n) = e^{-u-\lambda}\frac{(u+\lambda)^n}{n!}$
$P(X = k|X + Y = n) = n!/(k!(n-k)!)\frac{\lambda^k u^{n-k}}{(u+\lambda)^n} =$

Chapter 3: Expectation

3.1

After one trial: $E(X) = 2 * cP(w) + c/2 * P(L) = 2c * 1/2 + c/2 * 1/2 = c + c/4 = 5c/4$
https://an1lam.github.io/learning/2019-12-12-all-of-statistics-ch3-notes/

3.2

$V(X) = E(X - u_x)^2 = \sum_{x_i \in X}(x_i - u_X)^2 P(X = x_i) = 0$ since the terms are non-negative and probabilities have to sum to one, the only way is $x_i = u_X$, then $P(X = u_X) = 1$.

3.3

$E(Y_n) = E(max\{X_1, X_2, ..., X_n\}) = \int_0^1 Y_n f(x)dx = \int_0^1 max\{X_1, X_2, ..., X_n\}dx =$

3.4

After one jump: $E(X_1) = p * 1 + (1 - p) * (-1) = 2p - 1$ After n jump: $E(X_n) = \sum_{i=1}^n E(X_i) = n(2p - 1)$

$V(X_1) = p * 1 + (1 - p) * 1 - (2p - 1)^= 1 - (2p - 1)^2$ Since the jumps are i.i.d. $V(X_n) = \sum_{i=1}^n V(X_i) = nV(X_i) = n * (1 - (2p - 1)^2)$

3.5

$E(X) = 1 * p + 2 * (1 - p)p + 3 * (1 - p)^2 p + ... = \sum x * (1 - p)^{x-1} p = p \sum x * (1 - p)^{x-1} = -pd/dp(\sum(1 - p)^x) = p(-1/p)' = p * (-p^{-1})' = pp^{-2} = p/p^2 = 1/p$

3.6

????????????

3.7

$\int_0^\infty 1 - F(x)dx = x(1 - F(x))|_0^\infty - \int_0^\infty x - f(x)dx = \int_0^\infty xf(x)dx = E(X)$

3.8

$E(\bar{X}) = E(1/n\sum_{i=1}^{n} X_i) = 1/nE(\sum_{i=1}^{n} X_i) = 1/n\sum_{i=1}^{n} E(X_i) = n/nu = u$ by the linearity of the expected value.

$V(\bar{X}) = V(1/n\sum_{i=1}^{n} X_i) = 1/n^2V(\sum_{i=1}^{n} X_i) = 1/n^2\sum_{i=1}^{n} V(X_i) = n/n^2\sigma^2 = \sigma^2/n$ - by the linearity of variance when observations are i.i.d.

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

Where, $E(X_i - \bar{X})^2 = E(X_i^2 + \bar{X}^2 - 2X_i\bar{X}) = E(X_i^2) - u^2 + E(\bar{X}^2) - u^2 = V(X_i) - V(\bar{X}) = \sigma^2 - \sigma^2/n$

$V(\bar{X}) = E(\bar{X}^2) - (E\bar{X})^2$

$\frac{n}{n-1}(n\sigma^2 - \sigma^2)/n = \frac{n\sigma^2-\sigma^2}{n-1} = \sigma^2$

3.9

3.10

$E(Y) = E(e^x) = 1/\sqrt{2\pi}*\int_0^\infty e^x e^{-x^2/2}dx = 1/\sqrt{2\pi}*\int_0^\infty e^{x-x^2/2}dx = 1/\sqrt{2\pi}*\int_0^\infty e^{(2x-x^2)/2}dx = 1/\sqrt{2\pi} * \int_0^\infty e^{1/2-(x/\sqrt{2}-1/\sqrt{2})^2} = \sqrt{e}\sqrt{\pi}/\sqrt{2}erf(\frac{x-1}{\sqrt{2}})|_{-\infty}^{\infty} = \sqrt{2e\pi}$ By the same logic: $E(Y^2) = E(e^2x) = \sqrt{2\pi}e^2$ $Var(Y) = \sqrt{2\pi}e^2 - 2e\pi$

3.11

3.12

Bernoulli: $E(X) = 1 * p + 0 * (1 - p) = p$, $E(X^2) = p, V(X) = p - p^2 = p(1 - p) = pq$.

Poisson: $E(X) = \sum_{x=0}^{\infty} xe^{-\lambda}\lambda^x/x! = e^{-\lambda}\sum_{x=0}^{\infty}\lambda^x/(x-1)! = \lambda * e^{-\lambda}\sum_{x=0}^{\infty}\lambda^{x-1}/(x-1)! = \lambda * e^{-\lambda}e^{\lambda-1} = \lambda * e^{-1}$

Uniform: $E(X) = \int_a^b x/(b-a)dx = (a+b)/2, E(X^2) = \int_a^b x^2/(b-a)^2dx = a^3 - b^3/3(b-a)^2 = (a^2+ab+b^2)/3(b-a), V(X) = E(X^2) - E(X)$

Exponential: $E(X) = \int_0^\infty x\lambda e^{-\lambda x} = \lambda \inf_0^\infty xe^{-\lambda x} = xe^{-\lambda x}/(-\lambda)|_0^\infty - \int_0^\infty e^{-\lambda x}/(-\lambda)dx = 1/\lambda, E(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x} = 2/\lambda^2$-apply integration by parts two times.

Gamma: $E(X) = \int_0^\infty x\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{\beta x}dx = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}x^\alpha e^{\beta x}dx = \int_0^\infty \frac{\alpha*\beta^{\alpha+1}}{\beta*\Gamma(\alpha+1)}x^\alpha e^{\beta x}dx = \alpha/\beta.E(X^2) = \int_0^\infty x\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha+1}e^{\beta x}dx = \int_0^\infty \frac{\alpha(\alpha+1)*\beta^{\alpha+2}}{\beta^2\Gamma(\alpha+2)}x^{\alpha+1}e^{\beta x}dx = \frac{\alpha(\alpha+1)}{\beta^2}$

3.13

$E(X) = X * f_X(x|H)P(H) + X * f_X(x|T)P(T) = 1/2(\int_0^1 xdx + \int_3^4 xdx) = 1/2 * (1/2 + 7/2) = 1/2 * 8/2 = 8/4 = 2$ - by the thoerem of total expectation. The same way $E(X^2) = 1/2(\int_0^1 x^2dx + \int_3^4 x^2dx) = 19/3$ So, $V(X) = 19/3 - 4 = 7/3$.

3.14

We know that according to the theorem $3.19, Cov(X,Y) = E(XY) - E(X)E(Y)$.

$Cov(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j) = E(\sum_{i=1}^m a_i X_i \sum_{j=1}^n b_j Y_j) - E(\sum_{i=1}^m a_i X_i)E(\sum_{j=1}^n b_j Y_j) = E(\sum_{i=1}^m \sum_{j=1}^n a_i X_i b_j Y_j) - \sum_{i=1}^m a_i E(X_i) \sum_{j=1}^n b_j E(Y_j) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j E(X_i Y_j) - \sum_{j=1}^n \sum_{i=1}^m a_i b_j E(X_i)E($
$\sum_{i=1}^m \sum_{j=1}^n a_i b_j Cov(X_i, Y_j)$

3.15

$V(2X-3Y+8) = E(2X-3Y+8-E(2X-3Y+8))^2 = E(2X-3Y-E(2X-3Y))^2 = V(2X-3Y) = E((2X-3Y)^2) - E(2X-3Y) = E(4X^2+9Y^2-12XY) - 2E(X) + 3E(Y) = 4E(X^2) + 9E(Y^2) - 12E(XY) - 2E(X) + 3E(Y)$

3.16

Since $P(X=x, Y=y|X=x) = \frac{P(X=x,Y=y,X=x}{P(X=x)} = \frac{P(X=x,Y=y)}{P(X=x)} = P(Y=y|X=x)$
$E(r(X)s(Y)|X) = \int\int r(X)s(Y)f_{X,Y}(X=x, Y=y|X=x)dxdy = \int\int r(X)s(Y)f_Y(Y|X)dxdy = \int r(x)dx \int s(Y)f_Y(Y|X)dy == r(X)E(s(Y)|X)$

$E(r(X)|X) = \int r(X)f(X|X=x)dx = \int r(X)dx$?????????

3.17

$b(X) = E(Y|X=x), m = E(Y)$

$V(Y) = E(Y-m)^2 = E(y-b(X)+b(X)-m)^2 = E[(y-b(X))^2+(b(X)-m)^2+2*(y-b(X))(b(X)-m)] = E(V(X|Y)) + E(E(Y|X) - E(E(Y|X)))^2 + 2*E((y-b(X))(b(X)-m)) = E(V(X|Y)) + V(E(Y|X)) + 0$

3.18

$E(X|Y=y) = c$ for some constant c.

$E(XY) = \int\int xyf(x,y)dxdy = \int\int xyf(x|y)f(y)dxdy = \int yf(y)dy \int xf(x|y)dx = E(Y)c = cE(Y)$

$Cov(X,Y) = E(XY) - E(X)E(Y) = E(XY) - EE(X|Y)E(Y) = E(XY) - cE(Y) = cE(Y) - cE(Y) = 0$

3.19

3.20

3.21

$Cov(X,Y) = E(XY) - E(X)E(Y) = E(XY) - E(X)E(E(Y|X=x)) = E(XY) - E(X)^2$

$E(XY) = \int\int xyf(Y|X=x)f(X)dxdy = \int xf(X)dx \int yf(Y|X=x)dy = \int x^2 f(X)dx = E(X^2)$

3.22

0—-Y—-a—b–Z——-1 $P(Y=1\cap Z=0)=P(X\leq a)=a\neq P(Y=1)P(Z=0)=P(X\leq b)P(X\leq a)=ab$, therefore they are not independent.

$f(Y|Z)=P(Y\cap Z)/P(Z)$ $P(Z)=P(Z=0)+P(Z=1)=1$ $P(Y\cap Z)=P(Y=0\cap Z=0)+P(Y=1\cap Z=0)+P(Y=0\cap Z=1)+P(Y=1\cap Z=1)=0+a+1-b+b-a=1$ $E(Y|Z)=0*1+1*1=1$

Chapter 4: Inequalities

4.1

$u_X = 1/\lambda, \sigma_X = 1/\lambda^2$

$P(|X - u_X| \geq k\sigma_x) = P(X - u_X \geq k\sigma_x \cap -(X - u_X) \leq -k\sigma_x) = P(X - u_X \geq k\sigma_x) = 1 - e^(- \lambda k\sigma_x)$

$P(|X - u_X| \geq k\sigma_x) = Var(X)/(k\sigma_x)^2 = 1/(\lambda^2 * k * \sigma_x)^2$

4.2

$P(X \geq 2\lambda) = Var(X)/4\lambda^2 = \frac{1}{\sqrt{\lambda}\lambda^2} \leq \frac{1}{\lambda}$

4.3

$P(|\bar{X_n} - p| > \epsilon) \leq E(\bar{X_n} - p)^2/\epsilon^2$, where $E(\bar{X_n} - p)^2 = E(\bar{X_n}^2) + E(p^2) - 2E(\bar{X_n}p)) = (pq + p^2n^3)/n + p^2 - 2p^2 = (pq + p^2n^3)/n - p^2$

Assuming independence: $Var(\bar{X_n}^2) = 1/n^2 \sum_i Var(X_i) = npq/n^2 = pq/n = E(\bar{X_n}^2) - E(\bar{X_n})^2 = E(\bar{X_n}^2) - n^2p^2, E(\bar{X_n}^2) = pq/n + n^2p^2 = (pq + p^2n^3)/n$

$P(|\bar{X_n} - p| > \epsilon) \leq E(\bar{X_n} - p)^2/\epsilon^2 = \frac{(pq+p^2n^3)/n - p^2}{\epsilon^2}$

Also, $P(|\bar{X_n} - p| > \epsilon) \leq 2e^(- 2n\epsilon^2)$

4.4

I think, it should be $log\frac{a}{2}$ instead of $log\frac{2}{a}$

$P(|\bar{X_n} - p| > \epsilon) \leq 2e^{-2n|\frac{1}{2n}log\frac{a}{2}|} = 2e^{log\frac{a}{2}} = a$

$P(|\bar{X_n} - p| \leq \epsilon) \geq 1 - a$

4.5

Use error function. $P(|Z| > t) = 2 * P(Z > t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2}dx =?????$

4.6

4.7

Assuming independence: $P(|\bar{X}_n| > t) = P(|\sum_i X_i| > tn) = P(|nX_i| > tn) = P(|X_i| > t) \leq \sqrt{\frac{2}{\pi}}e^{-t^2/2}/t$

$P(|\bar{X}_n| > t) \leq Var(X_n)/t^2 = \sigma^2/(nt^2)$

Chapter 5: Convergence of Random Variables

5.1

5.2

$E(X_n - b)^2 = E(X_n - E(X) + E(X) - b)^2 = E(X_n - E(X))^2 + E(X - b)^2 = Var(X_n) + E(X_n - b)^2$

5.3

$E(\bar{X} - u)^2 = E(1/n \sum X_i - u)^2 = E(\bar{X_n}^2 + u^2 - 2u\bar{(X)}) = E(\bar{X_n}^2) + u^2 - 2\frac{u}{n}\sum E(X_i) = E(\bar{X_n}^2) + u^2 - 2u^2 = E(\bar{X_n}^2) - u^2 = \frac{n}{n^2}E(X_1^2) - E(X_1)^2 = Var(X_1)$???????

5.4

Fix some $\epsilon > 0$, then the probability that $P(|X_n| > \epsilon) =$

5.5

Here we will use the exercise 2. It is enough to show that: $E(1/n \sum_i X_i^2) = p$ and $V(1/n \sum_i X_i^2) = 1/nVar(X_i^2) = pq/n \to 0$ as $n \to \infty$. Therefore, it converges in quadratic mean, which also implies convergence in probability.

5.6

$P(\bar{X} \geq 68) = P(10 * \frac{\bar{X}-68}{2.6} \geq 10 * (68 - 68)/2.6) = P(Z \geq 0) = 1/2$

5.7

$P(|X_n| > \epsilon) \leq E(X_n)^2/\epsilon^2$, where $E(X_n)^2 = Var(X_n) - E(X_n)^2 = \lambda_n - \lambda_n^2 = 1/n - 1/n^2$, so $P(|X_n| > \epsilon) \leq E(X_n)^2/\epsilon^2 = \frac{1}{n\epsilon^2} - \frac{1}{n^2\epsilon^2} \to 0$, as $n \to \infty$.

b) According to the theorem 5.5 point f, any continuous function g of X, converges to $g(0)$.

5.8

$E(X_i) = 1, Var(X_i) = 1, n = 100$

$P(Y < 90) = P(\bar{X_n} < 90/n) = P(\sqrt{n} * \frac{\bar{X_n}-1}{1} < \sqrt{n} * (90/n - 1)/1) = P(Z < 10 * (-0.1)) = P(Z < -1)$

5.9

$P(|X_n - X| > \epsilon) = 1/n \to 0$, as $n \to \infty$. Therefore, it converges in distribution too.

$E(X - X_n)^2 = E(X^2) + E(X_n^2) - 2E(XX_n)????$

5.10

$P(|Z| > t) = P(|Z|^k > t^k) \leq E(|Z|^k)/t^k$ by Markov inequality.

5.11

5.12

$lim_{n\to\infty} F_n(x) = F(x)$ for all x such that F is continuos. $lim_{n\to\infty} F_n(x) = lim_{n\to\infty} P(X_n \leq x) = P(X \leq x) =$

5.13

$P(X_n \leq x) = P(n*min\{Z_1, Z_2, ...Z_n\} \leq x) = P(min\{Z_1, Z_2, ...Z_n\} \leq x/n) = 1 - P(Z_1 > x/n)P(Z_2 > x/n)...P(Z_n > x/n) = 1 - P(Z_1 > x/n)^n = 1 - (1 - P(Z_1 \leq x/n))^n = 1 - exp(nlog(1 - F(x/n))) = 1 - exp(\frac{log(1-F(x/n))}{1/n})$

5.14

Let $X_1, ..., X_n$ be $Uniform(0, 1)$. Let $Y_n = \bar{X_n}^2$. Find the limiting distribution of Y_n.

$P(Y_n \leq y) = P(\bar{X_n}^2 \leq y) = P(1/n^2(\sum_{i=1}^n X_i)^2 \leq y) = P((\sum_{i=1}^n X_i)^2 \leq yn^2)$

We know that $\sqrt{n}\frac{\bar{X_n}-u}{\sigma}$ has a standard normal limiting distribution. Then by the Delta method $Y_n = \bar{X_n}^2$ has a $N(u^2, (2u)^2\sigma^2/n)$ limiting distribution.

Chapter 6: Models, Statistical Inference and Learning

(Below I assume independence because in example 6.8 it didn't say anything about independence but it was assumed).

6.1

Bias: $E_\theta(\bar{\theta_n}) - \theta$. Here $\theta = \lambda$ and $\bar{\theta_n} = n^{-1}\sum_i X_i$, so $E(\frac{\sum_{i=1}^n X_i}{n}) = 1/n\sum_{i=1}^n E(X_i) = 1/n * nE(X_1) = \lambda$-unbiased. Standard Error (Assuming independence): $se = se(\bar{\theta_n}) = \sqrt{Var(\bar{\theta_n})}$, so $Var(\frac{\sum_{i=1}^n X_i}{n}) = 1/n^2 Var(\sum_{i=1}^n X_i) = 1/n^2\sum_{i=1}^n Var(X_i) = n/n^2 Var(X_1) = \lambda/n, se = \sqrt{\lambda/n}$. MSE: $E(\bar{\theta_n} - \theta)^2 = bias(\bar{\theta_n})^2 + Var_\theta(\bar{\theta_n}) = \lambda^2 + \lambda/n = \frac{\lambda(\lambda * n+1)}{n}$

6.2

Bias (Assuming independence): $E(\bar{\theta_n}) = E(max\{X_1, X_2, ..., X_n\})$. We know that if the r.v. are non-negative then $E(X) = \int_0^\infty P(X \geq x)dx = \int_0^\infty(1 - P(X \leq x))dx$. So, $E(max\{X_1, X_2, ..., X_n\}) = \int_0^\theta(1 - P(max\{X_1, X_2, ..., X_n\} \leq x))dx = \int_0^\theta(1 - P(X_1 \leq x)P(X_2 \leq x)...P(X_n \leq x))dx = \int_0^\theta(1 - x^n)dx = \theta - \theta^{n+1}/(n+1)$-is biased. Standard Error: $E(max\{X_1, X_2, ..., X_n\}^2) = \int_0^\theta(1 - P(max\{X_1, X_2, ..., X_n\}^2 \leq x))dx = \int_0^\theta(1 - P(max\{X_1, X_2, ..., X_n\} \leq \sqrt{x})) = \int_0^\theta(1 - P(X_1 \leq \sqrt{x})P(X_2 \leq \sqrt{x})...P(X_n \leq \sqrt{x}))dx = \int_0^\theta(1 - \sqrt{x}^n)dx = \theta - \frac{\theta^{\frac{n}{2}+1}}{\frac{n}{2}+1}$, so $V(max\{X_1, X_2, ..., X_n\}^2) = \theta - \frac{\theta^{\frac{n}{2}+1}}{\frac{n}{2}+1} - (\theta - \theta^{n+1}/(n+1))^2$
MSE: $E(\bar{\theta_n} - \theta)^2 = bias(\bar{\theta_n})^2 + Var_\theta(\bar{\theta_n})$

6.3

Bias: $E_\theta(\bar{\theta_n}) = E(2 * \frac{\sum_{i=1}^n X_i}{n}) = 2n/nE(X_1) = 2 * \theta/2 = \theta$-unbiased. Standard error(Assuming independence): $Var(2 * \frac{\sum_{i=1}^n X_i}{n}) = 4n/n^2 Var(X_1) = 4/n * 1/12\theta^2 = 1/3n\theta^2, se = \sqrt{1/3n\theta^2}$ MSE: $E(\bar{\theta_n} - \theta)^2 = \theta^2 + 1/3n\theta^2$.

Chapter 7: Estimating the cdf and Statistical Functionals

7.1

$E(F_n\bar(x)) = E(1/n \sum_{i=1}^{n} I(X_i \leq x)$-is the expectation of the Bernouli with the probability of success $p = P(X_i \leq x) = F(x)$, therefore the $E(F_n\bar(X)) = F(x)$.

$Var(F_n\bar(x)$-by the same analogy it is the variance of the Bernouli with the probability of success $p = P(X_i \leq x) = F(x)$, thereofre $Var(F_n\bar(x) = \frac{F(x)(1-F(x)}{n}$,

$MSE : E(F_n\bar(x) - F(x))^2 = bias(F_n\bar(x))^2 + se(F_n\bar(x))^2 = \frac{F(x)(1-F(x)}{n}$,

Since $F_n\bar(x) \rightarrow^{q.m} F(x)$ then it converges in probability too.

7.2

The plug in estimate for p will be $1/n \sum_{i=1}^{n} X_i$, then the plug-in standard error is $(1/n \sum_{i=1}^{n} X_i^2) - \bar{X}^2$.

The plug in estimate for $p - q$ will be $\frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{m} \sum_{i=1}^{m} Y_i$. The plug-in standard error is $V(p-q) = V\bar(p) + V\bar(q) = \frac{\bar{X}(1-\bar{X}}{n} + \frac{\bar{Y}(1-\bar{Y}}{m}$, $\bar{se} = \sqrt{V(p-q)}$, and the 90 percent confidence interval gives $z_{\alpha/2} = z_{0.1/2} = z_{0.05} = 1.64$, therefore the confidence interval will be $\frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{m} \sum_{i=1}^{m} Y_i + -1.64 * \bar{se}$

7.4

Fix x. Then $F_n\bar(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$ from Exercise 1, we know that $E(F_n\bar(x)) = F(x), Var(F_n\bar(x)) = \frac{F(x)(1-F(x)}{n}$. Then according to the CLT $\sqrt{n}\frac{F_n\bar(x)-F(x)}{\sqrt{F(x)(1-F(x)}} \rightarrow N(0,1)$.

7.5

$Cov(F_n\bar(x), F_n\bar(y)) = E(F_n\bar(x) - F(x))(F_n\bar(y) - F(y)) = E(F_n\bar(x)F_n\bar(y)) - E(F_n\bar(x))E(F_n\bar(y))$
https://stats.stackexchange.com/questions/385568/covariance-of-an-empirical-distribution-function-evaluated-at-different-points

7.6

$Var(\bar{\theta}) = Var(F_n\bar(b) - F_n\bar(a)) = E((F_n\bar(b) - F_n\bar(a))^2) - E(F_n\bar(b) - F_n\bar(a))^2$

$E((\bar{F_n}(b) - \bar{F_n}(a))^2) = E(\bar{F_n}(b)^2 + \bar{F_n}(a)^2 - 2 * \bar{F_n}(b)\bar{F_n}(a)) = E(\bar{F_n}(b)^2) + E(\bar{F_n}(a)^2) - 2E(\bar{F_n}(b)\bar{F_n}(a))$

$E(\bar{F_n}(b)^2) = \frac{F(b)(1-F(b)}{n} + F(b)^2$ $E(\bar{F_n}(a)^2) = \frac{F(a)(1-F(a)}{n} + F(a)^2$ $E(\bar{F_n}(b)\bar{F_n}(a)) =???$

$E(\bar{F_n}(b) - \bar{F_n}(a))^2 = (E(\bar{F_n}(b)) - E(\bar{F_n}(a)))^2 = E(\bar{F_n}(b))^2 + E(\bar{F_n}(a))^2 - 2 * E(\bar{F_n}(b))E(\bar{F_n}(a)) = F(a)^2 + F(b)^2 - 2F(a)F(b)$

7.8

$\bar{p_1} = 0.9, \bar{p_2} = 0.85, \bar{p_1} - \bar{p_2} = 0.05$, $Var(\bar{p_1} - \bar{p_2}) = Var(\bar{p_1}) + Var(\bar{p_2}) = 0.9 * (1 - 0.9)/100 + 0.85 * (1 - 0.85)/100 = 0.0021$ $se = \sqrt{0.0021}$. The 80 percent confidence interval has $z_{0.2/2} = z_{0.1}$ $0.05 + -z_{0.1} * \sqrt{0.0021}$

Chapter 8: The Bootstrap

8.4

Following the hint, let's try to put n balls into n buckets. We can use the idea of the stars and bars to get $(\frac{n+n-1}{n})$.

8.5

We know that $P(X* = x*|X_1, ..., X_n) = 1/n$. Moreover, $E(X*|X_1, ..., X_n) = \sum_{i=1}^{n} X_i P((X* = x * |X_1, ..., X_n) = \bar{X}$

$E(\bar{X_n}^*|X_1, ..., X_n) = 1/n \sum_{i=1}^{n} E(X_i^*|X_1, ..., X_n) = n/nE(X_i^*|X_1, ..., X_n) = \bar{X}$, and $EE(\bar{X_n}^*|X_1, ..., X_n) = E(\bar{X}) = E(X_1)$.

$Var(\bar{X_i^*}|X_1, ..., X_n) = 1/n^2 n * Var(X_i^*|X_1, ..., X_n)$

$Var(X_i^*|X_1, ..., X_n) =$

Chapter 9: Parametric Inference

9.1

$E(X) = ab, E(X^2) = ab^2 + a^2b^2$. So,

$$\hat{a}\hat{b} = \sum_{i=1}^{n} X_i, \hat{a}\hat{b}^2 + \hat{a}^2\hat{b}^2 = \sum_{i=1}^{n} X_i^2$$

$$\hat{b} = \sum_{i=1}^{n} X_i/(\hat{a})$$

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

9.2

$E(X) = (a+b)/2, E(X^2) = Var(X) + E(X)^2 = (b-a)^2/12 + (a+b)^2/4 = \frac{(b-a)^2+3(b+a)^2}{12} = \frac{4b^2+4a^2+4ab}{12} = \frac{b^2+a^2+ab}{12}$

$$(\hat{a} + \hat{b})/2 = \hat{X}$$

,

$$\frac{(\hat{b} - \hat{a})^2 + 3(\hat{b} + \hat{a})^2}{12} = \frac{(\hat{b} - \hat{a})^2}{12} + \frac{(\hat{b} + \hat{a})^2}{4} = \sum_{i=1}^{n} X_i^2$$

$$\hat{a} = 2\hat{X} - \hat{b}$$

$$\frac{(\hat{b} - \hat{a})^2}{12} = \frac{\hat{b}^2 + \hat{a}^2 - 2\hat{a}\hat{b}}{12} = \sum_{i=1}^{n} X_i^2 - (\hat{X})^2$$

$\hat{a}^2 = 4\hat{X}^2 + \hat{b}^2 - 4\hat{X}\hat{b}$ $\hat{a}\hat{b} = 2\hat{X}\hat{b} - \hat{b}^2$ so,

$$\frac{\hat{b}^2 + \hat{a}^2 - 2\hat{a}\hat{b}}{12} = \frac{\hat{b}^2 + 4\hat{X}^2 - 2\hat{X}\hat{b}}{12} = \sum_{i=1}^{n} X_i^2 - (\hat{X})^2$$

MLE: If $X_{min} > a, X_{max} < b, L(a,b) = (1/(b-a))^n$, otherwise $L(a,b) = 0$ So, $\hat{a} = min(X_1, ..., X_n), \hat{b} = max(X_1, ..., X_n)$

$\rho = \int xdF(x) = \inf x f_x dx = E(X) = (\hat{b} + \hat{a})/2$

MSE: $MSE(\bar{\rho}) = bias(\bar{\rho}) + Var(\bar{\rho}) = 0 + (b-a)^2/(12n)$

9.3

$P(X \leq \rho) = P(Z \leq \frac{\rho-u}{\sigma}) = \Phi(\frac{\rho-u}{\sigma}) = 0.95, \rho = \Phi^{-1}(0.95)\sigma + u$, therefore, the $MLE(\rho) = \Phi^{-1}(0.95)\hat{\sigma} + \hat{u}$, where $\hat{u} = \bar{X}, \hat{\sigma} = 1/n\sum_{i=1}^{n}(X_i - \bar{X})^2$

The standard error of ρ is

9.4

$P(|\hat{\theta}-\theta| > \epsilon) = P(\hat{\theta} < \theta - \epsilon) = P(X_1 < \theta - \epsilon)^n = (\theta - \epsilon/\theta)^n = (1-\epsilon/\theta)^n \to 0$ as $n \to \infty$.

9.5

The method of moments: $E(X) = \lambda$, $\hat{\lambda} = \bar{X}$

MLE: $L(\lambda) = \lambda^{\sum_{i=1}^{n} x_i}(e^{-n\lambda})/_{i=1}^{n} x_i!$

$l(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i ln(\lambda) - \sum_{i=1}^{n} ln(x_i!)$

$dl(\lambda)/d\lambda = -n + \sum_{i=1}^{n} x_i/\lambda = 0, \lambda = \bar{X}$

Fisher information: $-\lambda + X ln(\lambda) - ln(X!)$

$s(\lambda) = -1 + X/\lambda \; ds(\lambda)/d\lambda = -X/\lambda^2$ So, $I(\lambda) = E(X/\lambda^2) = 1/\lambda$

9.6

a) $\psi = P(Y_1 = 1) = P(X_1 \geq 0) = P(Z_1 \geq -\theta) \; L(\theta) = \prod_{i=1}^{n} exp(-(x-\theta)^2/2) = exp(\sum_{i=1}^{n}(-(x-\theta)^2/2)) \; l(\theta) = -1/2\sum_{i=1}^{n}(X_i-\theta)^2 \; dl(\theta)/d(\theta) = \sum_{i=1}^{n}(X_i-\theta) = 0, \hat{\theta} = \bar{X}$

$\hat{\psi} = P(Z_1 \geq -\hat{\theta}) = P(Z_1 \geq -\bar{X}) = 1 - \Phi(-\bar{X})$-by the Equivariance of the MLE.

b) The log of the distribution is $lnf(x) = -(x-\theta)^2/2$

The score function is $s(\theta) = (x - \theta)$

The Fisher information is $I(\theta) = E_\theta(-s'(\theta)) = E_\theta(1) = 1$

According to the Delta method we have: $\bar{X} + -2 * (1/\sqrt{I(\theta)})$ or $\bar{X} + -2$

c) $E(1/n\sum Y_i) = 1/n\sum_i E(Y_i) = P(X_1 > 0)$, therefore, by the q.m. it converges in probability to ψ. Hence, it is consistent.

d) $Var(\bar{\psi}) = Var(1/n\sum_i Y_i) = n/n^2 Var(Y_1) = P(X_1 > 0)(1 - P(X_1 > 0)/n = (1-\Phi(\theta))\Phi(\theta)/n$

By the delta method: $se(g(\theta)) = |g'(\hat{\theta})|se(\hat{\theta})$, where $g(\hat{\theta}) = 1 - \Phi(-\hat{\theta}) = 1 - (1 - \Phi(\hat{\theta})) = \Phi(\hat{\theta}), g'(\hat{\theta}) = \phi(\bar{X}) \; se(\hat{\theta}) = \sqrt{1/nVar(X_1)} = 1/\sqrt{n}$, therefore,

$(\sqrt{(1-\Phi(\theta))\Phi(\theta)/n})/(\phi(\bar{X})/\sqrt{n}) = \frac{\sqrt{(1-\Phi(\theta))\Phi(\theta)}}{\phi(\bar{X})}$

9.7

a) MLE for the binomial gives $\hat{p_1} = \bar{X_1} = p_1, \hat{p_2} = p_2$, therefore, $\hat{\psi} = \hat{p_1} - \hat{p_2}$

b) The distribution of X_1 can be rewritten as $p_1^{X_1}(1-p_1)^{1-X_1}$. The distribution of X_2 can be rewritten as $p_2^{X_2}(1-p_2)^{1-X_2}$. The likelyhood is $p_1^{X_1}p_2^{X_2}(1-p_1)^{n_1-X_1}(1-p_2)^{1-X_2}$. The log-likelyhood is $l = X_1 ln(p_1) + X_2 ln(p_2) + (1-X_1)ln(1-p_1) + (2-X_2)ln(1-p_2)$

$dl/dp_1^2 = [X_1/p_1 + (1-X_1)/(1-p_1)]' = -X_1/p_1^2 + (1-X_1)/(1-p_1)^2 \; dl/dp_1dp_2 = 0$
$dl/dp_1^2 = [X_2/p_2 + (2-X_2)/(1-p_2)]' = -X_2/p_2^2 + (2-X_2)/(1-p_2)^2$

The Fisher information is: $E(dl/dp_1^2) = E(-X_1/p_1^2 + (1 - X_1)/(1 - p_1)^2) = -\frac{1}{p_1} + frac1(1 - p_1)^2 + \frac{n_1 p_1}{(1-p_1)^2}$

$E(dl/dp_2^2) = E(-X_2/p_2^2 + (1 - X_2)/(1 - p_2)^2) = -\frac{1}{p_2} + frac1(1 - p_2)^2 + \frac{n_2 p_2}{(1-p_2)^2}$

c) $d\psi/dp_1 = 1, d\psi/dp_2 = -1, g = (1, -1)^T$

9.8

$L(u, \sigma) = \prod_{i=1}^n exp(-(X_i - u)^2/2\sigma) = \exp(\sum_{i=1}^n (-(X_i - u)^2/2\sigma))$

$l(u, \sigma) = -\sum_{i=1}^n \frac{(X_i - u)^2}{2\sigma}$

$dl/du^2 = [\sum_{i=1}^n \frac{(X_i - u)}{\sigma}]' = \frac{-n}{\sigma}$

$\frac{dl/dud\sigma = [\sum_{i=1}^n \frac{(X_i-u)}{\sigma}]' = -\sum_{i=1}^n \frac{(X_i-u)}{}}{\sigma^2}$

$\frac{dl/dud\sigma = [\sum_{i=1}^n \frac{(X_i-u)^2}{2\sigma^2}]' = -\sum_{i=1}^n \frac{(X_i-u)}{}}{\sigma^2}$

$\frac{dl/d\sigma^2 = [\sum_{i=1}^n \frac{(X_i-u)^2}{2\sigma^2}]' = -\sum_{i=1}^n \frac{(X_i-u)^2}{}}{\sigma^3}$

$E(dl/du^2) = \frac{-n}{\sigma}$

$E(dl/dud\sigma) = -1/\sigma^2 \sum_{i=1}^n E((X_i - u)) = 0 = E(dl/dud\sigma)$

$\frac{E(dl/d\sigma^2) = -E(\sum_{i=1}^n \frac{(X_i-u)^2}{}}{\sigma^3) = -1/\sigma^3 E(\sum_{i=1}^n (X_i-u)^2) = -n/\sigma^2}$

$dg/du = -\sigma/u^2, dg/d\sigma = 1/u$

Chapter 10: Hypothesis Testing and p-values

10.1

According to the definition the power $\beta(\theta^*) = P_{\theta^*}(X \in R)$, where R is the rejection region. In our case the $R = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$. Therefore, the probability $P(|X| \geq z_{\alpha/2}) = 1 - [P(X \leq z_{\alpha/2}) - P(X \leq -z_{\alpha/2})]$. ???????

10.2

10.3

Suppose we reject the H_0 while using the Wald's test. Then, $|W| > z_{\alpha/2}$ so, $\frac{\hat{\theta}-\theta_0}{\hat{}} > z_{\alpha/2}, -\theta_0 > \hat{} * z_{\alpha/2} - \hat{\theta}, \theta_0 < \hat{\theta} - \hat{} * z_{\alpha/2}$ and $-\frac{\hat{\theta}-\theta_0}{\hat{}} < -z_{\alpha/2}, \theta_0 < -\hat{} * z_{\alpha/2} + \hat{\theta}, \theta_0 < \hat{\theta} - \hat{} * z_{\alpha/2}$

Suppose, $\theta_0 \notin C$, then $\frac{\hat{\theta}-\theta_0}{\hat{se}} > z_{\alpha/2} > -z_{\alpha/2}$ and $-\frac{\hat{\theta}-\theta_0}{\hat{se}} < -z_{\alpha/2}$ Therefore, $|W| = |\frac{\hat{\theta}-\theta_0}{\hat{se}}| > z_{\alpha/2}$

10.4

10.5

a) $P_\theta(Y > c) = P_\theta(X_1 > c)P_\theta(X_2 > c)...P_\theta(X_n > c) = (1 - P_\theta(X_1 < c))^n = (1 - c/\theta)^n$ if $c < \theta$, and $P_\theta(Y > c) = 0$ otherwise.

b) $a = (1 - c/\theta)^n, a^{1/n} = 1 - c/\theta, c = (a^{1/n} + 1)/\theta$ where $a = 0.05, \theta = 1/2$.

c) $P(Y > 0.48) = 1 - \prod_{i=1}^{n} P(X_i < 0.48) = 1 - 0.96^{20} = 0.56$, there is no sufficient evidence to reject H_0.

d) $P(Y > 0.52) = 0$, since $X_i \leq \theta = 1/2$ under the H_0 for all $i \in [1, n]$.

10.6

Let's use the Wald test statistics. $\hat{p} = 922/1919 = 0.48, \hat{se} = \sqrt{0.48 * (1 - 0.48)/1919} = 0.011, W = \frac{0.48-0.5}{0.011} = -1.818$. Therefore, the p-value is equal to $P(|Z| > 1.818) = 0.07$, an evidence againsts H_0. A confidence interval is $0.48 + -0.011 * 2$

10.7

a) $\bar{T} = \frac{0.225+0.262+0.217+0.240+0.230+0.229+0.235+0.217}{8} = 0.231$

$\bar{S} = \frac{0.209+0.205+0.196+0.210+0.202+0.207+0.224+0.223+0.220+0.201}{10} = 0.209$

$V\hat{ar}(T) = \frac{(0.225-0.231)+(0.262-0.231)+(0.217-0.231)+(0.240-0.231)+(0.230-0.231)+(0.229-0.231)+(0.235-0.231)+(0.217-0.2}{10}$
$\frac{0.007}{8} = 0.000875$

$V\hat{ar}(S) = \frac{0.007}{10} = 0.0007$

$\hat{se} = \sqrt{0.0007 + 0.000875} = 0.039$

$Z = \frac{\bar{T,/.}-\bar{S}}{\hat{se}} = \frac{0.231-0.209}{0.039} = 0.564$

$P(|Z| > 0.564)$???

10.8

a) Under H_0 we have that X_i is $N(0,1)$ for all $i \in [1,n]$. Also we know that if X,Y is $N(0,1)$, then $X+Y$ has $N(0,2)$ distribution. Then, T has the $N(0,1/n)$ distribution. Therefore, $P(T > c) = P(\sqrt{n}T > \sqrt{n}c) = P(Z > \sqrt{n}c) = 1 - \Phi(\sqrt{n}c)$, so to have a size $\alpha = 1 - \Phi(\sqrt{n}c), \Phi(\sqrt{n}c) = 1 - \alpha, c = \Phi^{-1}(1-\alpha)/\sqrt{n}$

b) Under H_1 the distribution of T is $N(1,1/n)$, hence, $\beta(1) = P_\theta(T > c) = P_\theta(T-1 > c-1) = P_\theta(\sqrt{n}(T-1) > \sqrt{n}(c-1)) = P_\theta(Z > \sqrt{n}(c-1)) = 1 - \Phi(\sqrt{n}(c-1))$

10.9

Under H_1: according to the exercise 10.1 we have that $\beta(\theta_1) = \Phi(\frac{\theta_0-\theta_1}{\hat{se}} + z_{\alpha/2}) + \Phi(\frac{\theta_0-\theta_1}{\hat{se}} - z_{\alpha/2})$. As $n \to \infty$, $\hat{se} \to 0$, so, $\frac{\theta_0-\theta_1}{\hat{se}} \to -\infty$. Therefore, $\beta(\theta_1) \to 1$.

10.13

$L(u) = \prod_{i=1}^n exp(-(X_i - u)^2/2\sigma)$ $l(u) = -\sum_{i=1}^n (X_i - u)^2/2\sigma$

Therefore, by the likelihood test: $\lambda = 2log(\frac{L(\theta)}{L(\theta_0)} = 2log\frac{\prod_{i=1}^n exp(-(X_i-u)^2/2\sigma)}{\prod_{i=1}^n exp(-(X_i-u_0)^2/2\sigma)} = 2log\frac{\prod_{i=1}^n exp(-(X_i-u)^2}{\prod_{i=1}^n exp(-(X_i-u_0)^2} =$ $2log\frac{exp(-\sum_{i=1}^n (X_i-u)^2}{exp(-\sum_{i=1}^n (X_i-u_0)^2} = 2log(exp(-\sum_{i=1}^n (X_i-u)^2 + \sum_{i=1}^n (X_i-u_0)^2)) = 2log(exp(u_0^2 - u^2 - 2X_i * u_0 + 2X_i * u)) = 2(u_0^2 - u^2 - 2X_i * u_0 + 2X_i * u)$, so $l\hat{a}mbda = 2(u_0 - \bar{X})$, from the MLE.

By the Wald test: $dl(u)/du = \sum_{i=1}^n (X_i - u)/\sigma$

$I(u) = -E_u(-n/\sigma) = n/\sigma$

$se = 1/\sqrt{n/\sigma} = \sqrt{\sigma}/\sqrt{n}$

$W = \frac{\hat{u}-u_0}{\hat{se}} = \sqrt{n}\frac{\bar{X}-u_0}{\sqrt{\sigma}}$

10.15

$L(p) = \prod_{i=1}^n p^{X_i}(1-p)^{1-X_i} = p^{\sum_{i=1}^n X_i}(1-p)^{\sum_{i=1}^n (1-X_i)} = p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i)} = p^{\sum_{i=1}^n X_i}(1-p)^n(1-p)^{-\sum_{i=1}^n X_i)} = (\frac{p}{1-p})^{\sum_{i=1}^n X_i)}(1-p)^n$

$l(p) = n * ln(1-p) + \sum_{i=1}^n X_i * ln(\frac{p}{1-p})$

$\lambda = 2log\frac{(\frac{p}{1-p})^{\sum_{i=1}^n X_i}(1-p)^n}{(\frac{p_0}{1-p_0})^{\sum_{i=1}^n X_i)}(1-p_0)^n} = 2log\frac{(}{p(1-p_0)}p_0(1-p))^{\sum_{i=1}^n X_i}\frac{(}{1-p}1-p_0)^n = 2n\frac{1-p}{1-p_0} + 2\sum_{i=1}^n X_i * \frac{p(1-p_0)}{p_0(1-p)}$

$dl(p)/dp = \sum_{i=1}^n X_i/p - (n - \sum_{i=1}^n X_i)/(1-p)$

$I(p) = -E(-\sum_{i=1}^n X_i/p^2 - (n - \sum_{i=1}^n X_i)/(1-p)^2) = np/p^2 + n - np/(1-p)^2$

$se = \sqrt{1/I(p)}$

$W = \frac{\hat{p}-p_0}{se}$

Chapter 11: Bayesian Inference

11.1

Take the prior θ to have $N(a, b^2)$ and $X_1, ..., X_n$ to have $N(\theta, \sigma^2)$ distributions. $f(\theta|X^n) = \frac{f(X^n|\theta)*f(\theta)}{\int f(X^n|\theta)f(\theta)d\theta}$

$f(X^n|\theta) = L(\theta) = \prod_{i=1}^n \frac{1}{2\sigma\sqrt{2\pi}} exp(-\frac{(X_i-\theta)^2}{2\sigma^2}) = (\frac{1}{2\sigma\sqrt{2\pi}})^n \prod_{i=1}^n exp(-\frac{(X_i-\theta)^2}{2\sigma^2}) = (\frac{1}{2\sigma\sqrt{2\pi}})^n e^{(-\sum_{i=1}^n \frac{(X_i-\theta)^2}{2\sigma^2}}$

$f(\theta)L(\theta) = [\frac{1}{b\sqrt{2\pi}} e^{-(\theta-a)^2/2b^2}]*[(\frac{1}{2\sigma\sqrt{2\pi}})^n e^{(-\sum_{i=1}^n \frac{(X_i-\theta)^2}{2\sigma^2})}] = \frac{1}{2b\sigma^n\sqrt{2\pi}^{n+1}} e^{(-\sum_{i=1}^n \frac{(X_i-\theta)^2}{2\sigma^2})-(\theta-a)^2/2b^2}) = \frac{1}{2b\sigma^n\sqrt{2\pi}^{n+1}} e^{(\frac{-\sum_{i=1}^n (X_i-\theta)^2}{2\sigma^2} - \frac{(\theta-a)^2}{2b^2})}$???????

11.2

11.3

$f(\theta)f(X^n|\theta) = 1/\theta(1/\theta)^n = (1/\theta)^{n+1}$ if $max(X^n) \leq \theta$

$f(\theta)f(X^n|\theta) = 0$, otherwise.

$\int_0^\infty (1/\theta)^{n+1} d\theta$, has no solution if $max(X^n) \leq \theta$, otherwise is equal to c.

Therefore, $f(\theta|X^n)$ is undefined if $max(X^n) \leq \theta$, and otherwise is 0.

11.4

a) The MLE of $\rho = p_1 - p_2$ is $\hat{p_1} - \hat{p_2}$, where $\hat{p_1} = \bar{X} = 30/50 = 3/5, \hat{p_2} = \bar{Y} = 40/50 = 4/5$
Let $n = 30, m = 40, N = 50$.

$$L(p_1, p_2) = p_1^X (1-p_1)^{n-X} p_2^Y (1-p_2)^{m-Y}$$

$$l(p_1, p_2) = X \ln p_1 + (n-X) \ln 1 - p_1 + Y \ln p_2 + (m-Y) \ln 1 - p_2$$

$$\frac{dl(p_1, p_2)}{dp_1} = X/p_1 - (n-X)/(1-p_1)$$

$$\frac{dl(p_1, p_2)}{dp_1 p_2} = X/p_1 - (n-X)/(1-p_1) = 0$$

$$\frac{dl(p_1, p_2)}{dp_1^2} = -X/p_1^2 - (n-X)/(1-p_1)^2$$

Therefore, the Fisher matrix is: $E(X/p_1^2 + (1-X)/(1-p_)^2) = n/p_1 + 1/(1-p_1)^2 - np_1/(1-p_1)^2 = \frac{n(1-p_1)^2+p_1-np_1^2}{p_1(1-p_1)^2} = \frac{n+np_1^2-2np_1+p_1-np_1^2}{p_1(1-p_1)^2} = \frac{n-2np_1+p_1}{}$

11.6

a) $L(\lambda) = \prod_{i=1}^n \lambda^{X_i} e^{-\lambda}/X_i!$

$\prod_{i=1}^n \lambda^{X_i} e^{-\lambda}/X_i! * \beta^a/G(a)\lambda^{a-1}e^{-b\lambda} \propto [\lambda^S e^{-n*}] * [\lambda^{a-1}e^{-b\lambda}] \propto \lambda^{S+a-1}e^{-\lambda(n+b)}$

Where $S = \sum_{i=1}^n X_i$

Here the posterior is also a Gamma distribution with $G(S+a, n+b)$. Therefore, the posterior mean is equal to $\frac{S+a}{n+b}$

b) $l(\lambda) = S * ln(\lambda) - n\lambda - ln(X!)$ $\frac{dl(\lambda)}{d\lambda} = S/\lambda - n$ $\frac{dl(\lambda)}{d\lambda^2} = -S/\lambda^2$

$I_n(\lambda) = E(-\frac{dl(\lambda)}{d\lambda^2}) = E(S)/\lambda^2 = n/\lambda$ Therefore, Jeffrie's prior is $\sqrt{n/\lambda}$

So, the posterior will be $[\prod_{i=1}^n \lambda^{X_i}e^{-\lambda}/X_i!] * [\sqrt{n/\lambda}] \propto [\lambda^S e^{-n\lambda}] * [\sqrt{n/\lambda}]$

11.7

11.8

$P(H_0|X^n) = \frac{P(X^n|H_0)P(H_0)}{P(X^n|H_0)P(H_0)+P(X^n|H_1)P(H_1)}$

Where, $P(X^n|H_0) = \prod_{i=1}^n 1/\sqrt{2\pi}exp(-(X_i)^2/2) * 1/2$

$P(X^n|H_1) = \int f(X|u)f(u)du = \int (1/\sqrt{2\pi})^n * exp(-\sum(X_i-u)^2/2) * 1/b\sqrt{2\pi}exp(-(u)^2/2b^2)du \propto exp(-\sum(X_i-u)^2/2) * exp(-(X_i)^2/2b^2) =$

$-\sum(X_i-u)^2/2 - (u)^2/2b^2 = (-b^2S^2 - nb^2u^2 + 2b^uS - 2u^2)/2b^2$

Chapter 12: Statistical Decision Theory

12.1

We have $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ Then, the risk is defined as $R(\theta, \hat{\theta}) = \int L(\theta, \hat{\theta}) f(x, \theta) dx$. Therefore, the Bayesian risk is defined as $r(f, \hat{\theta}) = \int R(\theta, t\hat{h}ef(\theta) d\theta$.

a) $P(p|X) = \frac{P(X|p)P(p)}{P(X)}$ $P(X|p)P(p) \propto p^X(1-p)^{n-X}p^{a-1}(1-p)^{b-1} \propto p^{X+a-1}(1-p)^{n+b-X-1}$, which is itself $Beta(X+a, n+b-X)$. Therefore, the Bayes estimate of p is $\hat{p} = \frac{X+a}{a+n+b}$.

Since the loss function is the squared error, the risk is the $MSE = Var(\hat{p}) + bias^2(\hat{p})$.

$Var(\hat{p}) = Var(\frac{X+a}{a+n+b}) = \frac{Var(X)}{(a+n+b)^2} = \frac{np(1-p)}{(a+n+b)^2}$

$Bias(\hat{p}) = E(\hat{p}) - p = \frac{E(X)}{(a+n+b)} + \frac{a}{(a+n+b)} - p = \frac{np}{(a+n+b)} + \frac{a}{(a+n+b)} - p$.

http://people.stat.sfu.ca/ lockhart/richard/830/11$_3/lectures/bayesian_estimation/web.pdf$

Chapter 13: Linear and Logistic Regression

13.1

$RSS = \sum_{i=1}^{n}(Y_i - \hat{Y_i})^2 = \sum_{i=1}^{n}(Y_i - b_0 - X_i b_1)^2$ $dRSS/db_0 = 2\sum_{i=1}^{n}(Y_i - b_0 - X_i b_1) = 0$
$nb_0 = \sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i$ $b_0 = \bar{Y_n} - b_1 \bar{X_n}$

$dRSS/db_1 = 2\sum_{i=1}^{n} X_i(Y_i - b_0 - X_i b_1) = 0,$
$dRSS/db_1 = \sum_{i=1}^{n} X_i Y_i - b_0 \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$
$dRSS/db_1 = \sum_{i=1}^{n} X_i Y_i - \bar{Y_n} \sum_{i=1}^{n} X_i - b_1 \bar{X_n} \sum_{i=1}^{n} X_i - b_1 \sum_{i=1}^{n} X_i^2 = 0$
$dRSS/db_1 = \sum_{i=1}^{n} X_i Y_i - \bar{Y_n} \sum_{i=1}^{n} X_i = -b_1 \bar{X_n} \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^2$
$dRSS/db_1 = \sum_{i=1}^{n} X_i Y_i - \bar{Y_n} \sum_{i=1}^{n} X_i = b_1(\sum_{i=1}^{n} X_i^2 - \bar{X_n} \sum_{i=1}^{n} X_i)$
$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \bar{Y_n} \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2 - \bar{X_n} \sum_{i=1}^{n} X_i}$
$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{Y_n}\bar{X_i}}{\sum_{i=1}^{n} X_i^2 - n\bar{X_n}^2}$
$b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$
$E(\hat{\sigma^2}) = \frac{1}{n-2} E(\sum_{i=1}^{n} \hat{\epsilon_i}^2)$
$\hat{\epsilon_i} = Y_i - \hat{Y_i} = Y_i - \hat{b_0} - \hat{b_1} X_i$
$E(\hat{\epsilon_i}^2) = E(Y_i - \hat{b_0} - \hat{b_1} X_i)^2 = E(Y_i)^2 + E(\hat{b_0} - \hat{b_1} X_i)^2 - 2E(Y_i(\hat{b_0} - \hat{b_1} X_i)) = E(Y_i)^2 + E(\hat{b_0})^2 + E(\hat{b_1} X_1)^2 - 2E(\hat{b_0}\hat{b_1} X_i) - 2E(Y_i(\hat{b_0} - \hat{b_1} X_i))$
$E(\hat{b_1}) = E[\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}]$
$E(\hat{b_1}) = E[\frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}]$ from the second equation of the b_1. Than,
$E(\hat{b_1}) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})E(Y_i)}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(b_0 + b_1 X_i)}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$
Now use the fact that for any constant $c, \sum_{i=1}^{n}(X_i - \bar{X})c = c(\sum_{i=1}^{n} X_i - n\bar{X}) = 0$ and write
$E(\hat{b_1}) = \frac{b_1 \sum_{i=1}^{n}(X_i - \bar{X})X_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$
Also, note that $\sum_{i=1}^{n}(X_i - \bar{X})X_i = \sum_{i=1}^{n} X_i^2 - \bar{X}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i^2 - \bar{X}n\sum_{i=1}^{n} X_i/n = \sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \sum_{i=1}^{n} X_i^2 + \bar{X}^2 - 2\sum_{i=1}^{n} X_i * \bar{X} = \sum_{i=1}^{n} X_i^2 + \bar{X}^2 - 2\sum_{i=1}^{n} X_i^2/n$
Therefore, $E(\hat{b_1}) = b_1$
$E(\hat{b_0}) = E(\bar{Y} - \bar{X}\hat{b_1}) = \bar{Y} - \bar{X}b_1 = (\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i)/n = \sum_{i=1}^{n} b_0/n = b_0$
$Var(\hat{b_1}) = Var(\frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}) = (\frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2})^2 Var(Y_i)$
$Var(Y_i) = Var(b_0 + b_1 X_i + \epsilon_i) = \sigma^2$
$Var(\hat{b_1}) = \frac{(\sum_{i=1}^{n}(X_i - \bar{X}))^2\sigma^2}{(\sum_{i=1}^{n}(X_i - \bar{X})^2)^2} = (\frac{(\sum_{i=1}^{n}(X_i - \bar{X}))}{\sum_{i=1}^{n}(X_i - \bar{X})^2})^2\sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$
$Var(\hat{b_0}) = Var(\bar{Y} - \hat{b_1}\bar{X}) = \bar{X}^2 Var(\hat{b_1}) = \frac{\sum_{i=1}^{n} X_i^2}{n^2}(\frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2})$

$Cov(\hat{b_0},\hat{b_1}) = E(\hat{b_0}\hat{b_1}) - E(\hat{b_0})E(\hat{b_1})$
$E(\hat{b_0}\hat{b_1}) = E(\hat{b_0}\hat{b_1}) - E(\hat{b_0})E(\hat{b_1})$
$E(\hat{b_0}\hat{b_1}) = E((\bar{Y_n} - \hat{b_1}\bar{X_n})\hat{b_1}) = E(\bar{Y_n}\hat{b_1}) - E(\hat{b_1}^2\bar{X_n}) = E(\hat{Y_n}b_1) - \bar{X_n}E(\hat{b_1}^2)$
$E(\hat{b_1}^2) = Var(\hat{b_1}) + E(\hat{b_1})^2 = \frac{\sigma^2}{ns_X^2} + b_1^2$
$E(\bar{Y_n}b_1) = E(\hat{b_1}\sum_{i=1}^n Y_i/n) = \sum_{i=1}^n E(\hat{b_1}Y_i)/n$
$E(\hat{b_1}(b_0 + b_1X_i + \epsilon_i))) = b_0b_1 + b_1^2X_i$
$E(\bar{Y_n}b_1) = E(\hat{b_1}\sum_{i=1}^n Y_i/n) = \sum_{i=1}^n E(\hat{b_1}Y_i)/n = b_0b_1 + b_1^2\bar{X_n}$
$E(\hat{b_0}\hat{b_1}) = b_0b_1 + b_1^2\bar{X_n} - \bar{X_n}\frac{\sigma^2}{ns_X^2} - \bar{X_n}b_1^2$
$Cov(\hat{b_0},\hat{b_1}) = -\bar{X_n}\frac{\sigma^2}{ns_X^2}$

13.2

Look at the previous exercise.

13.3

$Y_i = bX_i + \epsilon$
$RSS = \sum_{i=1}^n(Y_i - \hat{Y_i})^2 = \sum_{i=1}^n(Y_i - \hat{b}X_i - \epsilon)^2$
$dRSS/db = -2\sum_{i=1}^n(Y_i - \hat{b}X_i)X_i = 0$
$\sum_{i=1}^n X_iY_i - \hat{b}\sum_{i=1}^n X_i^2 = 0$, $\hat{b} = \frac{\sum_{i=1}^n X_iY_i}{\sum_{i=1}^n X_i^2}$
$Var(\hat{b}) = Var(\frac{\sum_{i=1}^n X_iY_i}{\sum_{i=1}^n X_i^2}) = \frac{(\sum_{i=1}^n X_i^2}{)}(\sum_{i=1}^n X_i^2)^2 Var(Y_i) = \frac{\sigma^2}{\sum_{i=1}^n X_i^2}$
$se(\hat{b}) = \frac{\sigma}{\sqrt{\sum_{i=1}^n X_i^2}}$
$E(\hat{b}) = E(\frac{\sum_{i=1}^n X_iY_i}{\sum_{i=1}^n X_i^2}) = \sum_{i=1}^n \frac{X_i}{\sum_{i=1}^n X_i^2}E(Y_i) = \beta,$
Take $E(Y_i) = X_ib$, to get $E(\hat{b}) = \sum_{i=1}^n \frac{X_i}{\sum_{i=1}^n X_i^2}X_ib = \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i^2}b = b,$
Therefore, to get consitancy we assume that $E(Y_i|X_i) = X_ib$, $Var(Y_i|X_i) = Var(\epsilon|X_i) = \sigma$, Y_i are independent of each other and normally distributed.
If we assume that $Y_i|X_i$ is normally distributed with mean X_ib and variance σ^2 than the MLE will be
$\prod_{i=1}^n f(X_i,Y_i) \propto \prod_{i=1}^n f(Y_i|X_i) = \sigma^{-n}exp(\frac{(Y_i-u_i)^2}{2\sigma^2}) = -n\log\sigma - 1/2\sigma^2\sum_{i=1}^n(Y_i - bX_i)^2$.
Our estimate of b will be consistent.

13.4

$bias(\hat{R_{tr}}(S)) = E(\hat{R_{tr}}(S)) - R(S)$.
$R(S) = \sum_{i=1}^n E(Y_i\hat{(S)} - Y_i^*)^2 = \sum_{i=1}^n E(Y_i\hat{(S)}^2) + \sum_{i=1}^n E((Y_i^*)^2) - 2\sum_{i=1}^n E(Y_i\hat{(S)}Y_i^*)$
$E(\hat{R_{tr}}(S)) = \sum_{i=1}^n E((Y_i\hat{(S)} - Y_i)^2) = \sum_{i=1}^n E((Y_i\hat{(S)}^2) + \sum_{i=1}^n E(Y_i^2) - 2\sum_{i=1}^n E(Y_i(\hat{S})Y_i)$
$Y^* = X_ib, Y_i\hat{(S)} = X_i(S)b(S)$.
$E(\hat{R_{tr}}(S)) = \sum_{i=1}^n E((Y_i\hat{(S)} - Y_i)^2) = \sum_{i=1}^n E[((Y_i\hat{(S)} - X_ib) - (Y_i - X_ib))^2] = \sum_{i=1}^n E[((Y_i\hat{(S)} - X_ib)^2] + \sum_{i=1}^n E[(Y_i - X_ib))^2] - 2\sum_{i=1}^n E[(Y_i\hat{(S)} - X_ib)(Y_i - X_ib)]$
https://statweb.stanford.edu/ candes/teaching/stats300c/Lectures/Lecture19.pdf

13.5

$H_0 : b_1 = 17b_0, H_1 : b_1 \neq 17b_0$

Let $\sigma = b_1 - 17b_0$ and the null hypthesis be $\sigma = 0$. $Var(\sigma) = Var(b_1) + 289Var(b_0)$

$t = \frac{b_1 - 17b_0}{(\sqrt{Var(b_1) + 289Var(b_0)})}$

$\hat{t} = \frac{\hat{b_1} - 17\hat{b_0}}{(\sqrt{Var(\hat{b_1}) + 289Var(\hat{b_0})})}$

Chapter 14: Multivariate Models

14.1

Let a be a vector of length k and let X be a random vector of the same length with mean u and variance $\sum$. Then $E(a^T X) = E(a_1 X_1 + a_2 X_2 + ... + a_n X_n) = a_1 E(X_1) + a_2 E(X_2) + ... + a_n E(X_n) = a^T u$, where u is defined as $(E(X_1), E(X_2), ..., E(X_n))$ is the mean of a random vector by the definition 14.1.
The variance: $Var(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n \sum_{j=1}^n [E(a_i X_i a_j X_j) - E(a_i X_i) E(a_j X_j)] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j [E(X_i X_j)$
$E(X_i)E(X_j)]$

Moreover, $a^T \sigma a = a^T \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) a_j = \sum_{i=1}^n \sum_{j=1}^n a_i Cov(X_i, X_j) a_j$, so they are equal.

$E(AX) = \sum_{i=1}^k E(a_i^T X) = \sum_{i=1}^k a_i^T u = Au$, where a_i - is the i-th row of the matrix A.

14.2

The log-likelihood is $l(p) = \sum_{j=1}^k X_j \log p_j$ $l(p)/dp = \sum_{j=1}^k X_j / p_j$ $l(p)/dp^2 = -\sum_{j=1}^k X_j / {p_j}^2$

$-E(l(p)/dp^2) = E(\sum_{j=1}^k X_j / {p_j}^2) = E(\frac{X_1}{p_1} + \frac{X_2}{p_2} + ... + \frac{X_k}{p_k}) = E(p^T X) = n^2$

$\hat{se} = \sqrt{1/n^2} = 1/n.$

Chapter 15: Inference About Independence

15.1

$Y||Z$, then $P(Z = 0 \cap Y = 0) = P(Z = 0)P(Y = 0) = p_{0.}p_{.0} = p_{00}$ $P(Z = 1 \cap Y = 1) = P(Z = 1)P(Y = 1) = p_{1.}p_{.1} = p_{11}$ $P(Z = 1 \cap Y = 0) = P(Z = 1)P(Y = 0) = p_{1.}p_{.0} = p_{10}$ $P(Z = 0 \cap Y = 1) = P(Z = 0)P(Y = 1) = p_{0.}p_{.1} = p_{01}$. So, $\phi = \frac{p_{0.}p_{.0}p_{1.}p_{.1}}{p_{1.}p_{.0}p_{0.}p_{.1}} = 1$. Moreover, $log(\phi) = log(1) = 0$.

15.2

15.3

To compute the MLE of ϕ lets compute the MLE of $p_{00}, p_{01}, p_{10}, p_{11}$ and then use the equivariance of the MLE . Since we can represent $X = (X_{00}, X_{01}, X_{10}, X_{11})$ as a multinoimal with (n, p) where $p = (p_{00}, p_{01}, p_{10}, p_{11})$, than we would get the corresponding MLE to be $\hat{p}_{ij} = X_{ij}/n$. Therefore, $\hat{\phi} = \frac{X_{00}X_{11}/n^2}{X_{01}X_{10}/n^2} = \frac{X_{00}X_{11}}{X_{01}X_{10}}$. Moreover, using the equivaraince of the MLE we would get the MLE $\hat{\gamma} = \log \hat{\phi}$.

Chapter 16: Causal Inference

16.1

$E(Y|X = 1) = 0 + 1/2 = 0.5$ $E(Y|X = 0) = 0 + 0/2 = 0$
$E(C_1) = 1 + 0 + 0 + 0/4 = 0.25$ $E(C_0) = 1 + 0 + 0 + 1/4 = 0.5$

16.2

Since $C(x)$ and X are independent than $E(C(x)|X) = E(C(x))$. Keeping in mind that $Y = C(X)$ we can write $\theta(x) = E(C(x)) = E(C(x)|X = x) = E(Y|X = x) = r(x)$.

16.3

$\theta = E(C_1) - E(C_0) = E(C_1|X = 1)P(X = 1) - E(C_1|X = 0)P(X = 0) - E(C_0|X = 1)P(X = 1) - E(C_0|X = 0)P(X = 0) = E(C_1|X = 1)P(X = 1) - E(C_1|X = 0) + E(C_1|X = 0)P(X = 1) - E(C_0|X = 1)P(X = 1) - E(C_0|X = 0) + E(C_0|X = 0)P(X = 1) = E(Y|X = 1)P(X = 1) - E(C_1|X = 0) + E(C_1|X = 0)P(X = 1) - E(C_0|X = 1)P(X = 1) - E(Y|X = 0) + E(Y|X = 0)P(X = 1)$. Since $Y|X$ is binary with probability p, then

Chapter 17: Directed Graphs and Conditional Independence

17.1

We know from the conditional independance that $f(xy|z) = f(x,y,z)/f(z) = f(x|z)f(y|z)$

We also know from the definition that

$f(x,y|z) = f(x,y,z)/f(z) = f(x,y,z)/f(z) = f(x|y,z)f(y|z)$. Therefore, $f(x|z) = f(x|y,z)$. The two statements are equivallent.

17.2

$f(x,y|z) = f(x|z)f(y|z) = f(y|z)f(x|z) = f(y,x|z)$.

$f(x|y,z) = f(x|z)$, $f(h(X) \leq x|y,z) = f(X \leq h^{-1}(x)|y,z) = f(X \leq h^{-1}(x)|z) = f(h(X) \leq x|z)$, therefore, $f(h(x)|z) = f(h(x)|y,z)$. ????????????????????????

17.3

$f(x,y|z) = f(x,y,z)/f(z) = f(x|y,z)f(y|z)$ $Z = 0$: $f(x = 0, y = 0|z = 0) = 0.405/0.5 = 0.81$ $f(x = 0, y = 1|z = 0) = 0.045/0.5 = 0.09$ $f(x = 1, y = 0|z = 0) = 0.045/0.45 = 0.09$ $f(x = 1, y = 1|z = 0) = 0.005/0.5 = 0.01$

$Z = 1$: $f(x = 0, y = 0|z = 1) = 0.125/0.5 = 0.25$ $f(x = 0, y = 1|z = 1) = 0.125/0.5 = 0.25$ $f(x = 1, y = 0|z = 1) = 0.125/0.5 = 0.25$ $f(x = 1, y = 1|z = 1) = 0.125/0.5 = 0.25$

$f(x = 0|z = 0) = f(x = 0, z = 0)/f(z = 0) = f(x = 0|y = 0, z = 0) = f(x = 0, y = 0, z = 0)/f(y = 0, z = 0)$. The same for the rest can be checked.

$f(x = 0) = f(x = 0|z = 0)f(z = 0) + f(x = 0|z = 1)f(z = 1) = (0.45/0.5 + 0.25/0.5) * 0.5 = 0.7$ $f(x = 1) = f(x = 1|z = 0)f(z = 0) + f(x = 1|z = 1)f(z = 1) = (0.05/0.5 + 0.25/0.5) * 0.5 = 0.3$

$f(y = 0) = f(y = 0|z = 0)f(z = 0) + f(y = 0|z = 1)f(z = 1) = (0.45/0.5 + 0.25/0.5) * 0.5 = 0.7$ $f(y = 1) = f(y = 1|z = 0)f(z = 0) + f(y = 1|z = 1)f(z = 1) = (0.05/0.5 + 0.25/0.5) * 0.5 = 0.3$ by the symmetry.

$P(X = 0, Y = 0) = P(X = 0, Y = 0|Z = 0)P(Z = 0) + P(X = 0, Y = 0|Z = 1)P(Z = 1) = P(X = 0|Z = 0)P(Y = 0|Z = 0)P(Z = 0) + P(X = 0|Z = 1)P(Y = 0|Z = 1)P(Z = 1) = (0.45/0.5 * 0.45/0.5 + 0.25/0.5 * 0.25/0.5) * 0.5 = 0.53$ $\not P(X = 0)P(Y = 0) = 0.7 * 0.7 = 0.49$

17.4

In all of the cases we have a sitation when Y is not a collider. Therefore, according to the rules of the d-separation X, Z are d-separated given Y. Hence, by the theorem 17.10 we have that they are independent given Y.

17.5

X, Z are dependent given Y since they are d-connected since Y is a collider. Moreover, by the rules of the d-seperation X, Z are d-connected, therefore by the theorem 17.10 they are dependent marginally.

17.7

For all $i \in [1, 4]$ the Z_i and X are d-separated by the rules of the d-seperation since for each i there is a Y_i that is a collider for X and Z_i

$P(Z, Y, X) = P(Z|X, Y)P(Y|X)P(X)$

$P(Z|Y) = P(Z, Y)/P(Y) = \sum_x XP(X, Y, Z)/P(Y) = \sum_X P(Z|X, Y)P(Y|X)P(X)/P(Y) = \frac{\sum_X P(Z|X,Y)P(Y|X)P(X)}{P(Y|X)P(X)}$

$\sum_{x=0}^{1} P(Z|X = x, Y)P(Y|X = x)P(X = x) = \sum_{x=0}^{1} (\frac{e^{2(x+y)-2}}{1+e^{2(x+y)-2}})^x (\frac{1}{1+e^{2(x+y)-2}})^{1-x} (\frac{e^{4x-2}}{1+e^{4x-2}})^x (\frac{1}{1+e^{4x-2}})^{1-}$ $0.5)^{1-x} = \frac{1}{1+e^{2y-2}} \frac{1}{1+e^{-2}} 0.5 + \frac{e^{2(1+y)-2}}{1+e^{2(1+y)-2}} \frac{e^{-2}}{1+e^{-2}} 0.5$

$f(Z|Y := 1) = \sum_x f(x)f(z|x, y)$

Chapter 18: Undirected Graphs

18.1

a) $X_1 - X_2 - X_3$ b) $X_1, X_2 - X_3$ c) X_1, X_2, X_3

18.2

a) $X_1 - X_2 - X_3 - X_4$ b) $X_1 - X_4, X_2, X_3 - X_4$ c) $X_1 - X_2 - X_3 - X_4 - X_1$