Homework 6

Question 1

Let X be a binomially distributed random variable. For $i = 0, 1, \ldots, n$, $P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$ for some probability p between 0 and 1. Suppose that Y is Poisson distributed with mean X. Find the conditional expectation $E[X|Y]$.
Solution: When $X = 0$, the random variable Y is equal to 0 with probability 1. When $X = i$, the probability of the event that Y equals k is

$$P(Y = k | X = i) = \frac{i^k}{k!} e^{-i} \tag{1}$$

for $k = 0, 1, 2, 3, 4, \ldots$

Consider the conditioning on the event $Y = 0$. For $i = 0, 1, \ldots, n$, the probability of $X = i$ and $Y = 0$ is

$$P(X = i, Y = 0) = \begin{cases} \binom{n}{i} p^i (1-p)^{n-i} & \text{if } i = 0, \\ \binom{n}{i} p^i (1-p)^{n-i} \cdot \frac{i^0}{0!} e^{-i} & \text{if } i \geq 1. \end{cases} \tag{2}$$

(We avoid the ambiguity of 0^0 by distinguishing two cases.)

The expectation of X given $Y = 0$ is

$$\frac{1}{\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} e^{-i}} \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} e^{-i} \tag{3}$$

Using identities

$$(1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i$$
$$nx(1+x)^{n-1} = \sum_{i=0}^{n} \binom{n}{i} i x^i,$$

we can simplify the answer to

$$\frac{np/e}{(1 - p + p/e)} \tag{4}$$

For $k > 0$, the expectation of X given $Y = k$ is

$$\frac{1}{\sum_{i=1}^{n}\binom{n}{i}p^i(1-p)^{n-i}\frac{i^k}{k!}e^{-i}}\sum_{i=1}^{n} i\binom{n}{i}p^i(1-p)^{n-i}\frac{i^k}{k!}e^{-i}$$
$$= \frac{\sum_{i=1}^{n}\binom{n}{i}i^{k+1}K^i}{\sum_{i=1}^{n}\binom{n}{i}i^k K^i}$$

where K represents the constant $\frac{p}{(1-p)e}$.

Question 2

We flip a biased coin, which is head with probability p, and tail with probability $1-p$. If the coin turns up head, let Y be a Gaussian distributed random variable with mean 1 and variance 1. Otherwise, if the coin turns up to be tail, let Y be Gaussian with mean 0 and variance 1. Find the conditional probability of head given that the value of random variable Y is y, for $y \in \mathbb{R}$.
Naive solution: Apply the mixed version of Bayes theorem. Denote the conditional pdf of random variable Y given the event of seeing a head by

$$f_{Y|H}(y) = \frac{1}{\sqrt{2\pi}}e^{-(y-1)^2/2}, \tag{5}$$

and the the conditional pdf of random variable Y given the coin toss is tail by

$$f_{Y|T}(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}. \tag{6}$$

The conditional probability of head given $Y = y$ is

$$\begin{aligned} P(H|Y=y) &= \frac{f_{Y|H}(y)p}{f_{Y|T}(y)(1-p) + f_{Y|H}(y)p} \\ &= \frac{p\exp(-(y-1)^2/2)}{(1-p)\exp(-y^2/2) + p\exp(-(y-1)^2/2)}. \end{aligned}$$

Solution according to the abstract definition of conditional expectation. We can model the experiment by a product probability space $\Omega = \{H, T\} \times \mathbb{R}$. A probability measure P is defined in a way that

(i) event $\{(x, y) : x = H, y \in [a, b]\}$ has probability

$$p\int_a^b \frac{1}{\sqrt{2\pi}}e^{-(t-1)^2/2}\, dt. \tag{7}$$

(ii) event $\{(x, y) : x = T, y \in [a, b]\}$ has probability

$$(1-p)\int_a^b \frac{1}{\sqrt{2\pi}}e^{-t^2/2}\, dt. \tag{8}$$

We check that the function

$$g(y) = \frac{f_{Y|H}(y)p}{f_{Y|T}(y)(1-p) + f_{Y|H}(y)p} \tag{9}$$

is indeed the solution. Firstly $g(y)$ is a function of y, and hence is $\sigma(Y)$-measurable.

Let $\mathbf{1}_H(x,y)$ denote the indicator function of the event that the coin toss result is head, i.e.,

$$\mathbf{1}_H(x,y) = \begin{cases} 1 & \text{if } x = H \\ 0 & \text{if } x = Y. \end{cases} \tag{10}$$

It remains to check that

$$\int_B g(y)\ dP(x,y) = \int_B \mathbf{1}_H(x,y)\ dP(x,y) \tag{11}$$

for all $\sigma(Y)$-measurable sets. By the $\pi - \lambda$ theorem, and the fact that closed intervals generate the Borel algebra on $\mathbb{R}$, it suffices to check it for B in the form $\{H,T\} \times [a,b]$.

$$\begin{aligned}
R.H.S. &= \int_{\{H,T\}\times[a,b]} \mathbf{1}_H(x,y)\ dP(x,y) \\
&= \int_{\{H\}\times[a,b]} \mathbf{1}_H(x,y)\ dP(x,y) \\
&= \int_a^b p f_{Y|H}(y)\ dy \\
&= \int_a^b g(y)(f_{Y|T}(y)(1-p) + f_{Y|H}(y)p)\ dy \\
&= \int_a^b g(y)\ f_{Y|T}(y)(1-p)\ dy + \int_a^b g(y)\ f_{Y|H}(y)p\ dy \\
&= \int_{\{T\}\times[a,b]} g(y)\ dP(x,y) + \int_{\{H\}\times[a,b]} g(y)\ dP(x,y) \\
&= \int_{\{H,T\}\times[a,b]} g(y)\ dP(x,y) = L.H.S.
\end{aligned}$$

Question 3

We pick a random point in unit square with vertices $(0,0)$, $(0,1)$, $(1,1)$, and $(1,0)$. Let X and Y be the x- and y-coordinates, respectively. If the random point lies within the inscribed circle, let Z be equal to 1, otherwise let Z be 0. Compute the conditional probability distribution of the Y given X and Z. Justify your answer.

Solution: The random variable Z takes two values. We shall only consider conditioning on the event that $Z = 1$ below. The case for $Z = 0$ is analogous.

Let μ denote the measure function on the unit square. Let's denote the inscribed circle $(x - 0.5)^2 + (y - 0.5)^2 \le 0.25$ by C. The area of C is $\mu(C)$. Conditioned on the event $Z = 1$, the probability that a random point falls in a measurable set E is given by

$$\frac{1}{\mu(C)} \int_E \mathbf{1}_C \; d\mu. \tag{12}$$

We use the short-hand notation

$$\begin{aligned} f(x) &= 0.5 - \sqrt{0.25 - (x - 0.5)^2} \\ g(x) &= 0.5 + \sqrt{0.25 - (x - 0.5)^2} \end{aligned}$$

to represent the lower semi-circle and the upper semi-circle, respectively.

Conditioned on $Z = 1$, the x-coordinate is a continuous variable with density

$$p_{X|Z=1}(x) = (g(x) - f(x))/\mu(C). \tag{13}$$

Suppose $0 < a < b < 1$. The conditional probability $P(a \le X \le b|Z = 1)$ is equal to

$$\int_a^b p_{X|Z=1}(x) \; dx. \tag{14}$$

If we can find a function $h(y, x)$ such that

$$\frac{1}{\mu(C)} \int_R \mathbf{1}_C d\mu = \int_a^b \int_c^d h(y, x) p_{X|Z=1}(x) \; dy \; dx, \tag{15}$$

for all a, b, c and d satisfying $0 < a < b < 1, 0 < c < d < 1$, where R is the rectangle $\{(x, y) : a \le x \le b, c \le y \le d\}$, then we can say that $h(y, x)$ is the conditional probability distribution function of Y given $X = x$ and $Z = 1$.

To this end, we write $\frac{1}{\mu(C)} \int_R \mathbf{1}_C d\mu$ as a double integral

$$\begin{aligned} \frac{1}{\mu(C)} \int_a^b \int_{f(x)\vee c}^{g(x)\wedge d} \; dydx &= \int_a^b \frac{g(x) - f(x)}{\mu(C)} \int_{f(x)\vee c}^{g(x)\wedge d} \frac{1}{g(x) - f(x)} \; dydx \\ &= \int_a^b p_{X|Z=1}(x) \int_c^d \frac{\mathbf{1}_{[f(x),g(x)]}(y)}{g(x) - f(x)} \; dydx. \end{aligned}$$

In the last line, $\mathbf{1}_{[f(x),g(x)]}$ is the indicator function of the interval $[f(x), g(x)]$.

We can write the conditional pdf as

$$p_{Y|X=x,Z=1}(y) = \begin{cases} \frac{1}{g(x)-f(x)} & \text{if } y \in [f(x), g(x)], \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

This is the pdf of the uniform distribution between $f(x)$ and $g(x)$.

Question 4

Compute the characteristic function of
(i) an exponentially distributed random variables with mean $1/\lambda$.
Solution: This is a continuous-type random variable with pdf $\lambda e^{-\lambda x}\mathbf{1}_{x\geq 0}$.
The characteristic function is

$$\begin{aligned}\int_0^\infty e^{itx}\lambda e^{-\lambda x}\ dx &= \int_0^\infty \lambda e^{(it-\lambda)x}\ dx \\ &= \Big[\frac{\lambda}{it-\lambda}\Big]_0^\infty \\ &= \frac{\lambda}{\lambda - it}.\end{aligned}$$

`https://en.wikipedia.org/wiki/Exponential_distribution`
(ii) a geometrically distributed random variable with mean $1/p$.
Solution: This is a discrete-type random variable with pmf

$$P(X=n) = (1-p)^{n-1}p \tag{17}$$

for n=1,2,3,...
The characteristic function is

$$\begin{aligned}\sum_{n=1}^\infty e^{itn}(1-p)^{n-1}p &= pe^{it}\sum_{n=1}^\infty ((1-p)e^{it})^{n-1} \\ &= \frac{pe^{it}}{1-(1-p)e^{it}}.\end{aligned}$$

`https://en.wikipedia.org/wiki/Geometric_distribution`

Question 5

(Azuma's inequality, Klenke ex. 9.2.4) Show the following.
(i) If X is a random variable with $|X| \leq 1$ a.s., then there is a random variable Y with values in $\{-1,+1\}$ and with $E[Y|X] = X$.
Solution: Given a realization of the random variable $X = x$, we define the conditional probability of Y given $X = x$ by

$$P(Y=1) = (1+x)/2, P(Y=-1) = (1-x)/2. \tag{18}$$

Since X is between bounded between -1 and 1, both $(1+x)/2$ and $(1-x)/2$ are between 0 and 1. The conditional expectation of Y given X is

$$\frac{1+X}{2}\cdot 1 + \frac{1-X}{2}\cdot -1 = X. \tag{19}$$

The above information is sufficient for constructing a product probability space on which X and Y are defined.
(ii) For X as in (i) with $E[X] = 0$, infer that (using Jensen's inequality)

$$E[e^{\lambda X}] \le \cosh(\lambda) \le e^{\lambda^2/2} \tag{20}$$

for all $\lambda \in \mathbb{R}$.
Solution: The function $e^{\lambda x}$ is a convex function of x, taken values e^{λ} at $x = 1$ and $e^{-\lambda}$ at $x = -1$. We can upper bound the function $e^{\lambda x}$ by a linear function

$$\frac{1-x}{2}e^{-\lambda} + \frac{1+x}{2}e^{\lambda} \tag{21}$$

for $-1 \le x \le 1$.

By the monotone property of integral, we obtain

$$E[e^{\lambda X}] \le E[\frac{1-X}{2}e^{-\lambda} + \frac{1+X}{2}e^{\lambda}] = \frac{1}{2}e^{-\lambda} + \frac{1}{2}e^{\lambda} = \cosh(\lambda). \tag{22}$$

(We have used the assumption that $E[X] = 0$ in this step.)

The second inequality can be seen by comparing the power series expansions

$$\begin{aligned} \cosh(\lambda) &= 1 + \frac{1}{2!}\lambda^2 + \frac{1}{4!}\lambda^4 + \frac{1}{6!}\lambda^6 + \frac{1}{8!}\lambda^8 + \cdots \\ e^{\lambda^2/2} &= 1 + \frac{\lambda^2}{2} + \frac{1}{2!}\frac{\lambda^4}{4} + \frac{1}{3!}\frac{\lambda^6}{2^3} + \frac{1}{4!}\frac{\lambda^8}{2^4} + \cdots \end{aligned}$$

(iii) If $(M_n)_{n\in\mathbb{N}_0}$ is a martingale with $M_0 = 0$ and if there is a sequence $(c_k)_{k\in\mathbb{N}}$ of nonnegative numbers with $|M_n - M_{n-1}| \le c_n$ a.s. for all $n \in \mathbb{N}$, then

$$E[e^{\lambda M_n}] \le \exp(\frac{1}{2}\lambda^2 \sum_{k=1}^{n} c_k^2). \tag{23}$$

Solution: The argument in part (ii) can be modified if X has zero mean and is between $\pm c$ with probability 1,

$$E[e^{\lambda X}] \le \exp(\frac{1}{2}\lambda^2 c^2). \tag{24}$$

We proceed by induction. For $n = 0$, we have $E[e^{0\lambda}] = 1$. Using the convention that empty summation is 0, $\exp(\frac{1}{2}\lambda^2 \sum_{k=1}^{0} c_k^2) = 1$ as well.

For the induction step, we write $E[e^{\lambda M_n}]$ as an iterated conditional expectation,

$$E[E[e^{\lambda(M_n - M_{n-1}) + \lambda M_{n-1}} | \, M_1, M_2, \dots, M_{n-1}]] \tag{25}$$

The inner conditional expectation is equal to

$$\begin{aligned} &E[e^{\lambda(M_n - M_{n-1}) + \lambda M_{n-1}} | \, M_1, M_2, \dots, M_{n-1}] \\ &= e^{\lambda M_{n-1}} \cdot E[e^{\lambda(M_n - M_{n-1})} | \, M_1, M_2, \dots, M_{n-1}] \end{aligned}$$

The difference $M_n - M_{n-1}$ is between $-c_n$ and c_n w.p.1. By the martingale property, the conditional expectation of $M_n - M_{n-1}$ given M_1 to M_{n-1} is zero. Hence

$$E[e^{\lambda M_n} | \ M_1, M_2, \dots, M_{n-1}] \le e^{\lambda M_{n-1}} e^{\lambda^2 c_n^2/2}. \tag{26}$$

Taking expectation of both sides, we get

$$E[e^{\lambda M_n}] \le E[e^{\lambda M_{n-1}}] e^{\lambda^2 c_n^2/2}. \tag{27}$$

(iv) Under the assumptions of (iii), Azuma's inequality holds:

$$P(|M_n| \ge a) \le 2 \exp\Big(- \frac{a^2}{2(c_1^2 + \cdots + c_n^2)} \Big) \tag{28}$$

for all $a \ge 0$.

Solution: The final part is a standard application of Markov inequality. We first consider the positive part. Given any $a \ge 0$, the following equality

$$P(M_n \ge a) = P(e^{\lambda M_n} \ge e^{\lambda a}) \tag{29}$$

holds for any $\lambda > 0$.

By Markov inequality,

$$P(e^{\lambda M_n} \ge e^{\lambda a}) \le E[e^{\lambda M_n}]/e^{\lambda a} = e^{\frac{1}{2}\lambda^2 (c_1^2 + \cdots + c_n^2) - \lambda a}. \tag{30}$$

The exponent is a quadratic function, taking minimum value at

$$\lambda^* = \frac{a}{c_1^2 + \cdots + c_n^2}. \tag{31}$$

By picking $\lambda = \lambda^*$, we get

$$P(M_n \ge a) \le \exp(-\frac{a^2}{2(c_1^2 + \cdots + c_n^2)}). \tag{32}$$

We include the negative part by multiplying the right side by 2,

$$P(|M_n| \ge a) \le 2\exp(-\frac{a^2}{2(c_1^2 + \cdots + c_n^2)}). \tag{33}$$

Question 6

Question 6. (Polya's urn model) Consider an urn containing B black balls and W white balls initially. We carry out an iterative procedure described as follows. In each step, we draw a ball uniformly at random from the urn. If the ball is white, we return two white balls to the urn. If the ball is black, we return two black balls to the urn.

Let X_n be a binary random variable that equals 1 if the n-th ball drawn from the urn is in black color, and equal to 0 otherwise. Let S_n be the sum

$X_1 + X_2 + \cdots + X_n$. Thus, S_n is the number of black ball drawings in steps 1 to n.
(a) Prove that, for $0 \le k \le n$,

$$P(S_n = k) = \binom{n}{k} \frac{B(B+1)(B+2)\cdots(B+k-1)\cdot W(W+1)(W+2)\cdots(W+n-k-1)}{N(N+1)(N+2)\cdots(N+n-1)} \tag{34}$$

where $N = B + W$ is the initially number of balls in the urn.
Solution: Suppose that the black ball drawings occur at time $i_1, i_2, \ldots, i_k$. The probability of drawing a white ball at times j, for 1 to $i_1 - 1$, is $(W + j - 1)/(N + j - 1)$. Drawing a black balls at time i_1 has probability $B/(N + i_1 - 1)$. For $j = i_1 + 1, \ldots, i_2 - 1$, the probability of drawing a white ball at time j is $(W + j - 2)/(N + j - 1)$. Next, drawing a black ball at time i_2 has probability $(B + 1)/(N + i_2 - 1)$. Continue similarly, the probability of drawing black balls at time $i_1, \ldots, i_k$ has probability

$$\frac{B(B+1)(B+2)\cdots(B+k-1)\cdot W(W+1)(W+2)\cdots(W+n-k-1)}{N(N+1)(N+2)\cdots(N+n-1)\cdot}. \tag{35}$$

Since this is independent of $i_1, \ldots, i_k$, we multiply it by $\binom{n}{k}$ to obtain the answer.

In terms of Gamma function, the probability that $S_n = k$ is

$$\frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \cdot \frac{\Gamma(k+B)}{\Gamma(B)} \cdot \frac{\Gamma(n-k+B)}{\Gamma(W)} \cdot \frac{\Gamma(N)}{\Gamma(n+N)}. \tag{36}$$

(b) Show that S_n/n converges in distribution to the beta distribution with parameters B and W.
Solution: The probability $P(S_n \le \lfloor nx \rfloor)$ is equal to

$$\sum_{k=0}^{\lfloor nx \rfloor} \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \cdot \frac{\Gamma(k+B)}{\Gamma(B)} \cdot \frac{\Gamma(n-k+B)}{\Gamma(W)} \cdot \frac{\Gamma(N)}{\Gamma(n+N)}, \tag{37}$$

which can be re-written as

$$\frac{\Gamma(N)}{\Gamma(B)\Gamma(W)} \sum_{k=0}^{\lfloor nx \rfloor} \frac{\Gamma(n+1)}{\Gamma(n+N)} \cdot \frac{\Gamma(k+B)}{\Gamma(k+1)} \cdot \frac{\Gamma(n-k+W)}{\Gamma(n-k+1)}. \tag{38}$$

Use the asymptotic result

$$\frac{\Gamma(z+a)}{\Gamma(z+b)} \to z^{a-b} \tag{39}$$

for fixed a and b, as $z \to \infty$, to approximate the k-th term in the summation by

$$\frac{k^{B-1}}{n^{B-1}} \frac{(n-k)^{W-1}}{n^{W-1}} \frac{1}{n}. \tag{40}$$

Finally, take n to approach infinity and approximate the summation by Riemann integral.