Probability & Statistics for EECS: Final Exam Solution

Problem 1

(5 points) Please describe the pros and cons of *Bayesian statistical inference* and *Classical statistical inference*. Then explain why conjugate priors are important for Bayesian statistical inference.

Solution

Pros of Bayesian statistical inference:

- It provides a natural and principled way of combining prior information with data, within a solid decision theoretical framework. You can incorporate past information about a parameter and form a prior distribution for future analysis. When new observations become available, the previous posterior distribution can be used as a prior. All inferences logically follow from Bayes' theorem.
- It can calculate the probability that a hypothesis is true, which is generally what the researchers actually want to know.

Cons of Bayesian statistical inference:

- It requires you to know or construct a prior, but it does not tell you how to select the prior. If you do not proceed with caution, you can generate misleading results.
- It can produce posterior distributions that are heavily influenced by the priors.
- It may be computationally intensive due to integration over many parameters.

Pros of classical statistical inference:

- It tends to be less computationally intensive than the Bayesian statistical inference.

Cons of classical statistical inference:

- It does not take into account priors.
- It has only one well-defined hypothesis.

As mentioned above, both the selection of probabilistic prior and the computation to get the posterior are in general hard in practice. The existence of conjugate priors smoothen these processes of selection and computation via a set of well-behaved priors and posteriors, where the priors have good physical meanings for the parameters to estimate (*e.g.*, Beta prior is used for $[0, 1]$ random variables and Gamma prior is used for $[0, \infty)$ random variables, *etc.*) and the posteriors are efficiently computed via conjugacy (*e.g.*, Beta-Binomial, Gamma-Poisson, Normal-Normal, *etc.*).

Problem 2

(10 points) Let X and Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} cx^2y & \text{if } 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) (5 points) Find the value of constant c.

(b) (5 points) Find the conditional probability $P(Y \le \frac{X}{4} \big| Y \le \frac{X}{2})$.

Solution

(a) Since

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = 1,$$

we have

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx &= \int_0^1\int_0^x cx^2ydydx \\ &= \int_0^1 \left(\frac{cx^2y^2}{2}\right)\Big|_{y=0}^{y=x} dx \\ &= \int_0^1 \frac{cx^4}{2}dx \\ &= \frac{c}{10}.\end{aligned}$$

Therefore, we get $c = 10$.

(b)

$$\begin{aligned} P\left(Y \le \frac{X}{4}\Big| Y \le \frac{X}{2}\right) &= \frac{P\left(Y \le \frac{X}{4}\right)}{P\left(Y \le \frac{X}{2}\right)} \\ &= \int_0^1\int_0^{x/4} 10x^2ydydx \Big/ \int_0^1\int_0^{x/2} 10x^2ydydx \\ &= \int_0^1 5x^2y^2\Big|_{y=0}^{y=\frac{x}{4}} dx \Big/ \int_0^1 5x^2y^2\Big|_{y=0}^{y=\frac{x}{2}} dx \\ &= \frac{1}{16} \times 4 \\ &= \frac{1}{4}.\end{aligned}$$

Problem 3

(15 points) Let X and Y be two integer random variables with joint PMF

$$P_{X,Y}(x,y) = \begin{cases} \frac{1}{6\cdot 2^{\min(x,y)}} & \text{if } x, y \geq 0, |x-y| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) (5 points) Find the marginal distributions of X and Y. Are X and Y independent?

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

(c) (5 points) Find $E[X|Y = 2]$ and $\mathrm{Var}[X|Y = 2]$.

Solution

(a) Note that, we have

$$\begin{aligned} P_X(0) = &\ P_{X,Y}(0,0) + P_{X,Y}(0,1) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}, \\ P_X(1) = &\ P_{X,Y}(1,0) + P_{X,Y}(1,1) + P_{X,Y}(1,2) = \frac{1}{6}\left(1 + \frac{1}{2} + \frac{1}{2}\right) = \frac{1}{3}, \\ P_X(2) = &\ P_{X,Y}(2,1) + P_{X,Y}(2,2) + P_{X,Y}(2,3) = \frac{1}{6}\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{4}\right) = \frac{1}{6}, \\ P_X(3) = &\ P_{X,Y}(3,2) + P_{X,Y}(3,3) + P_{X,Y}(3,4) = \frac{1}{6}\left(\frac{1}{4} + \frac{1}{8} + \frac{1}{8}\right) = \frac{1}{12}. \end{aligned}$$

This case is similar for $P_Y(y)$. Therefore, in general, we have

$$P_X(k) = P_Y(k) = \begin{cases} \frac{1}{3} & \text{if } k = 0, \\ \frac{1}{3\cdot 2^{k-1}} & \text{if } k = 1, 2, 3, \dots \\ 0 & \text{otherwise.} \end{cases}$$

And they are not independent with example $P_X(0)P_Y(0) = \frac{1}{9} \neq \frac{1}{6} = P_{X,Y}(0,0)$.

(b)

$$\begin{aligned} P(X = Y) = &\ \sum_{k=0}^{\infty} P_{X,Y}(k,k) \\ = &\ \sum_{k=0}^{\infty} \frac{1}{6\cdot 2^k} \\ = &\ \frac{1}{3}. \end{aligned}$$

(c) We need to first find

$$P_{X|Y}(x|2) = \frac{P_{X,Y}(x,2)}{P(Y=2)} = 6P_{X,Y}(x,2),$$

so we have

$$P_{X|Y}(x|2) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{4} & \text{if } x = 2, 3 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, we have

$$\begin{aligned}
E[X|Y=2] =&\ 1\cdot\frac{1}{2}+2\cdot\frac{1}{4}+3\cdot\frac{1}{4}=\frac{7}{4},\\
E[X^2|Y=2] =&\ 1\cdot\frac{1}{2}+4\cdot\frac{1}{4}+9\cdot\frac{1}{4}=\frac{15}{4},\\
\Rightarrow \mathrm{Var}[X|Y=2] =&\ E[X^2|Y=2]-\big(E[X|Y=2]\big)^2=\frac{15}{4}-\frac{49}{16}=\frac{11}{16}.
\end{aligned}$$

Problem 4

(20 points) Let Z_1, Z_2 be two *i.i.d.* random variables satisfying standard normal distributions, *i.e.*, $Z_1, Z_2 \sim \mathcal{N}(0,1)$. Define

$$X = \sigma_X Z_1 + \mu_X;$$
$$Y = \sigma_Y \left(\rho Z_1 + \sqrt{1-\rho^2} Z_2\right) + \mu_Y,$$

where $\sigma_X > 0$, $\sigma_Y > 0$, $-1 < \rho < 1$.

(a) (5 points) Show that X and Y are bivariate normal.

(b) (5 points) Find the correlation coefficient between X and Y, *i.e.*, $\text{Corr}(X,Y)$.

(c) (5 points) Find the joint PDF of X and Y.

(d) (5 points) Find $E[Y|X]$ and $\text{Var}[Y|X]$.

Solution

(a) X and Y are bivariate normal, since every linear combination of X and Y,

$$t_1 X + t_2 Y = t_1(\sigma_X Z_1 + \mu_X) + t_2(\sigma_Y \left(\rho Z_1 + \sqrt{1-\rho^2} Z_2\right) + \mu_Y),$$

can also be written as a linear combination of Z_1 and Z_2,

$$(t_1\sigma_X + t_2\sigma_Y\rho)Z_1 + t_2\sigma_Y\sqrt{1-\rho^2}Z_2 + (t_1\mu_X + t_2\mu_Y), \quad \forall t_1, t_2,$$

which proves the bivariate normal distribution.

(b) By the property of covariance and the independence of Z_1, Z_2,

$$\begin{aligned}
\text{Cov}(X,Y) &= \text{Cov}\left(\sigma_X Z_1 + \mu_X, \sigma_Y(\rho Z_1 + \sqrt{1-\rho^2}Z_2) + \mu_Y\right) \\
&= \text{Cov}(\sigma_X Z_1, \sigma_Y \rho Z_1) + \text{Cov}(\sigma_X Z_1, \sigma_Y\sqrt{1-\rho^2}Z_2) \\
&= \sigma_X\sigma_Y\rho Var(Z_1) + \sigma_X\sigma_Y\sqrt{1-\rho^2}\text{Cov}(Z_1, Z_2) \\
&= \sigma_X\sigma_Y\rho
\end{aligned}$$

To get the correlation, we also need the variance of X and Y, which can be computed as

$$\begin{aligned}
\text{Var}(X) &= \text{Var}(\sigma_X Z_1 + \mu_X) = \sigma_X^2 \text{Var}(Z_1) = \sigma_X^2, \\
\text{Var}(Y) &= \text{Var}(\sigma_Y(\rho Z_1 + \sqrt{1-\rho^2}Z_2) + \mu_Y) = \sigma_Y^2\rho^2\text{Var}(Z_1) + \sigma_Y^2(1-\rho^2)\text{Var}(Z_2) = \sigma_Y^2.
\end{aligned}$$

Therefore,

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \rho.$$

(c) The inverse transformation is

$$\begin{aligned}
Z_1 &= \frac{X-\mu_X}{\sigma_X}, \\
Z_2 &= -\frac{\rho}{\sigma_X\sqrt{1-\rho^2}}(X-\mu_X) + \frac{1}{\sigma_Y\sqrt{1-\rho^2}}(Y-\mu_Y).
\end{aligned}$$

The Jacobian is

$$\frac{\partial(z_1, z_2)}{\partial(x,y)} = \begin{bmatrix} \frac{1}{\sigma_X} & 0 \\ -\frac{\rho}{\sigma_X\sqrt{1-\rho^2}} & \frac{1}{\sigma_Y\sqrt{1-\rho^2}} \end{bmatrix},$$

which has absolute determinant $\frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}}$.

So by the change of variables formula,

$$
\begin{aligned}
f_{X,Y}(x,y) =& f_{Z_1,Z_2}(z_1,z_2)\left|\frac{\partial(z_1,z_2)}{\partial(x,y)}\right| \\
=& \frac{1}{2\pi}\frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\Big(-\frac{1}{2}(z_1^2+z_2^2)\Big) \\
=& \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\Big((\frac{x-\mu_X}{\sigma_X})^2+(-\frac{\rho}{\sigma_X\sqrt{1-\rho^2}}(x-\mu_X)+\frac{1}{\sigma_Y\sqrt{1-\rho^2}}(y-\mu_Y))^2\Big)\right] \\
=& \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\Big((\frac{x-\mu_X}{\sigma_X})^2+(\frac{y-\mu_Y}{\sigma_Y})^2-2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y}\Big)\right].
\end{aligned}
$$

(d) When conditioning on $X=x$, by the linearity of expectation and independence of Z_1,Z_2, we have

$$
\begin{aligned}
E(Y|X=x) =& E\Big(\sigma_Y\rho Z_1+\sigma_Y\sqrt{1-\rho^2}Z_2+\mu_Y\big|\sigma_X Z_1+\mu_X=x\Big) \\
=& \sigma_Y\rho E\Big(Z_1\big|Z_1=\frac{x-\mu_X}{\sigma_X}\Big)+\sigma_Y\sqrt{1-\rho^2}E\Big(Z_2\big|Z_1=\frac{x-\mu_X}{\sigma_X}\Big)+\mu_Y \\
=& \sigma_Y\rho\frac{x-\mu_X}{\sigma_X}+\sigma_Y\sqrt{1-\rho^2}E(Z_2)+\mu_Y \\
=& \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)+\mu_Y.
\end{aligned}
$$

Therefore,

$$E(Y|X)=\rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X)+\mu_Y.$$

For the variance,

$$
\begin{aligned}
\mathrm{Var}(Y|X=x) =& \mathrm{Var}\Big(\sigma_Y\rho Z_1+\sigma_Y\sqrt{1-\rho^2}Z_2+\mu_Y\big|\sigma_X Z_1+\mu_X=x\Big) \\
=& \mathrm{Var}\Big(\sigma_Y\sqrt{1-\rho^2}Z_2+\sigma_Y\rho\frac{x-\mu_X}{\sigma_X}+\mu_Y\big|Z_1=\frac{x-\mu_X}{\sigma_X}\Big).
\end{aligned}
$$

Since adding a constant does not affect variance,

$$
\begin{aligned}
\mathrm{Var}(Y|X=x) =& \mathrm{Var}\Big(\sigma_Y\sqrt{1-\rho^2}Z_2\big|Z_1=\frac{x-\mu_X}{\sigma_X}\Big) \\
=& \sigma_Y^2(1-\rho^2)\mathrm{Var}(Z_2) \\
=& \sigma_Y^2(1-\rho^2).
\end{aligned}
$$

Therefore,

$$\mathrm{Var}(Y|X)=\sigma_Y^2(1-\rho^2).$$

Problem 5

(10 points) Let the random variable $X \sim \mathcal{N}(\mu, \tau^2)$. Given $X = x$, random variables $Y_1, Y_2, \ldots, Y_n$ are *i.i.d.* and have the same conditional distribution, *i.e.*, $Y_i|X = x \sim \mathcal{N}(x, \sigma^2)$. Define the sample mean $\bar{Y}$ as follows:

$$\bar{Y} = \frac{Y_1 + \cdots + Y_n}{n}.$$

(a) (4 points) Find the posterior PDF of X given $\bar{Y}$.

(b) (3 points) Find the MAP (Maximum a Posterior Probability) estimates of X given $\bar{Y}$.

(c) (3 points) Find the MMSE estimates of X given $\bar{Y}$. (We know that the MMSE of X given Y is given by $g(Y) = E[X|Y]$).

Solution

(a) Since $Y_i|X = x \sim \mathcal{N}(x, \sigma^2)$ for $i = \{1, 2, \cdots, n\}$, we can obtain that

$$\bar{Y} = \frac{Y_1 + \cdots + Y_n}{n}\Big|X = x \sim \mathcal{N}(x, \frac{\sigma^2}{n}).$$

Consider the Normal-Normal conjugacy. After observing $\bar{Y} = \bar{y}$, we can update our prior uncertainty for X using Bayes' rule,

$$f_{X|\bar{Y}}(x|\bar{y}) \propto f_{\bar{Y}|X}(\bar{y}|x) f_X(x) \propto e^{-\frac{n}{2\sigma^2}(\bar{y}-x)^2} e^{-\frac{1}{2\tau^2}(x-\mu)^2}.$$

Since we have a quadratic function of X in the exponent, we recognize the posterior PDF of X as a Normal PDF, *i.e.*, by completing the square,

$$\begin{aligned} e^{-\frac{n}{2\sigma^2}(\bar{y}-x)^2} e^{-\frac{1}{2\tau^2}(x-\mu)^2} &= \exp\Big[-\frac{1}{2}\Big(\frac{1}{\sigma^2\tau^2}(n\tau^2 + \sigma^2)x^2 + n\tau^2\bar{y}^2 + \sigma^2\mu^2 - 2n\tau^2 x\bar{y} - 2\sigma^2 x\mu\Big)\Big] \\ &\propto \exp\Big[-\frac{1}{2} \cdot \frac{n\tau^2 + \sigma^2}{\tau^2\sigma^2}\Big(x^2 - \frac{2n\tau^2}{n\tau^2 + \sigma^2}x\bar{y} + \frac{2\sigma^2}{n\tau^2 + \sigma^2}x\mu\Big)\Big] \\ &\propto \exp\Big[-\frac{1}{2} \cdot \frac{n\tau^2 + \sigma^2}{\tau^2\sigma^2}\Big(x - \frac{n\tau^2}{n\tau^2 + \sigma^2}\bar{y} - \frac{\sigma^2}{n\tau^2 + \sigma^2}\mu\Big)^2\Big]. \end{aligned}$$

We can obtain the posterior distribution of X given $\bar{Y}$ is

$$X|\bar{Y} \sim \mathcal{N}\left(\frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\bar{y} + \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\mu, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right).$$

(b) Rewrite the mean and variance in the posterior distribution as

$$\begin{aligned} \mu_1 &= \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\bar{y} + \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\mu, \\ \sigma_1^2 &= \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}. \end{aligned}$$

Since $f_{X|\bar{Y}}(x|\bar{y}) \propto \exp\big(-\frac{1}{2\sigma_1^2}(x - \mu_1)^2\big)$, we have

$$\begin{aligned} \hat{x}_{\text{map}} &= \arg\max_x f_{X|\bar{Y}=\bar{y}}(x) \\ &= \arg\max_x \exp\big(-\frac{1}{2\sigma_1^2}(x - \mu_1)^2\big) \\ &= \arg\min_x \frac{1}{2\sigma_1^2}(x - \mu_1)^2. \end{aligned}$$

Problem 5 continued on next page...

By the property of quadratic function, we can obtain that $\frac{1}{2\sigma_1^2}(x-\mu_1)^2$ reach its minimum at $x=\mu_1$, which means $\hat{x}_{\text{map}}=\mu_1$. Thus, the MAP estimator of X givn $\bar{Y}$ is

$$\hat{X}_{\text{map}} = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\bar{Y} + \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\mu.$$

(c) Since from (a) we have

$$X|\bar{Y} \sim \mathcal{N}\left(\frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\bar{y} + \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\mu, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right),$$

the MMSE of X given $\bar{Y}$ is

$$E(X|\bar{Y}) = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\bar{Y} + \frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\mu.$$

Problem 6

(10 points) Let $X_1 \sim \text{Expo}(\lambda_1)$, $X_2 \sim \text{Expo}(\lambda_2)$ and $X_3 \sim \text{Expo}(\lambda_3)$ be independent.

(a) (5 points) Find $E(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3)$ in terms of $\lambda_1, \lambda_2, \lambda_3$.

(b) (5 points) Find $P(X_1 = \min(X_1, X_2, X_3))$.

Solution

(a) Since X_1, X_2 and X_3 are independent, we have

$$\begin{aligned} & E\left(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3\right) \\ = \; & E\left(X_1 | X_1 > 1, X_2 > 2, X_3 > 3\right) + E\left(X_2 | X_1 > 1, X_2 > 2, X_3 > 3\right) + E\left(X_3 | X_1 > 1, X_2 > 2, X_3 > 3\right) \\ = \; & E\left(X_1 | X_1 > 1\right) + E\left(X_2 | X_2 > 2\right) + E\left(X_3 | X_3 > 3\right). \end{aligned}$$

According to the memoryless property of exponential distribution, for each $i \in \{1, 2, 3\}$, we have

$$E\left(X_i | X_i > i\right) = E\left(X_i\right) + i = \frac{1}{\lambda_i} + i.$$

Then it follows that

$$E\left(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3\right) = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + 6.$$

(b) Note that we have

$$P\left(X_1 = \min\left(X_1, X_2, X_3\right)\right) = P\left(X_1 \leq \min\left(X_2, X_3\right)\right).$$

Since $\min\left(X_2, X_3\right) \sim \text{Expo}\left(\lambda_2 + \lambda_3\right)$, then by the property of exponential distribution, we have

$$P\left(X_1 = \min\left(X_1, X_2, X_3\right)\right) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}.$$

Problem 7

(10 points) Let $X \sim \text{Expo}(\lambda)$, $Y \sim \text{Expo}(\lambda)$; X and Y are independent.

(a) (5 points) Find $E(X|X+Y)$.

(b) (5 points) Find $E(X^2|X+Y)$.

Solution

(a) By symmetry,

$$E\left(X|X+Y\right) = E\left(Y|X+Y\right),$$

and by linearity,

$$E\left(X|X+Y\right) + E\left(Y|X+Y\right) = E\left(X+Y|X+Y\right) = X+Y.$$

Therefore,

$$E\left(X|X+Y\right) = \frac{1}{2}\left(X+Y\right).$$

(b) Method 1: Since $X|X+Y \sim \text{Unif}(0, X+Y)$, we have

$$E\left(X^2|X+Y\right) = \int_0^{X+Y} x^2 \cdot \frac{1}{X+Y} dx = \frac{\left(X+Y\right)^2}{3}.$$

Method 2: For any constant $u > 0$, we have

$$E\left(X^2|X+Y=u\right) = \int_0^u x^2 f_{X|X+Y}\left(x|u\right) dx.$$

Either by symmetry or by Bayes' rule, we have $f_{X|X+Y}\left(x|u\right) = \frac{1}{u}$, then it follows that

$$E\left(X^2|X+Y=u\right) = \int_0^u x^2 \frac{1}{u} dx = \frac{u^2}{3}.$$

Therefore, we have

$$E\left(X^2|X+Y\right) = \frac{\left(X+Y\right)^2}{3}.$$

Problem 8

(10 points) Instead of predicting a single value for the parameter, we give an interval that is likely to contain the parameter: A $1-\delta$ confidence interval for a parameter p is an interval $[\hat{p}-\epsilon, \hat{p}+\epsilon]$ such that $P\left(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]\right) \geq 1-\delta$. Now we toss a coin with probability p landing heads and probability $1-p$ landing tails. The parameter p is unknown and we need to estimate its value from experiment results. We toss such coin N times. Let $X_i = 1$ if the ith result is head, otherwise 0. We estimate p by using $\hat{p} = \frac{X_1+\cdots+X_N}{N}$. Find the $1-\delta$ confidence interval for p, then discuss the impacts of δ and N. **Hint**: You can use the following Hoeffding bound: Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $E(X_i) = \mu$, $a \leq X_i \leq b$ for each $i = 1, \ldots, n$, where a, b are constants. Then for any $\epsilon \geq 0$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right).$$

Solution

Since $X_i \sim \mathrm{Bern}(p)$, then $E(X_i) = p$. Furthermore, there are only two possible values of X_i, 0 and 1, so $0 \leq X_i \leq 1$. Then by Hoeffding bound, we have

$$P(|\hat{p}-p| \geq \epsilon) = P\left(\left|\frac{1}{N}\sum_{i=1}^{N} X_i - p\right| \geq \epsilon\right) \leq 2e^{-2N\epsilon^2}.$$

Let $2e^{-2N\epsilon^2} = \delta$, we can get

$$\epsilon = \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}.$$

It follows that

$$\begin{aligned} & P(|\hat{p}-p| \geq \epsilon) \leq \delta \\ \Longleftrightarrow\ & P(|\hat{p}-p| < \epsilon) > 1-\delta \\ \Longleftrightarrow\ & P(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]) < \epsilon) > 1-\delta, \end{aligned}$$

where $\hat{p} = \frac{X_1+\cdots+X_N}{N}$, $\epsilon = \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}$.

Then we have found the confidence interval for p. It means that $\forall \delta > 0$, $|\hat{p}-p| < \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}$ with probability greater than $1-\delta$. We can see that if N rises, δ rises, then ϵ will drop, which means we can get a more accurate estimation for p.

Problem 9

(10 points) Show the following inequalities.

(a) (5 points) Let $X \sim \text{Pois}(\lambda)$. If there exists a constant $a > \lambda$, then

$$P(X \geq a) \leq \frac{e^{-\lambda}(e\lambda)^a}{a^a}.$$

(b) (5 points) Let X be a random variable with finite variance σ^2. Then for any constant $a > 0$,

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$

Solution

(a) Using Chernoff's technique and Markov's inequality, we have $\forall t > 0$,

$$P(X \geq a) = P\left(e^{tX} \geq e^{ta}\right) \leq \frac{E\left[e^{tX}\right]}{e^{ta}}.$$

By $X \sim \text{Pois}(\lambda)$, we have

$$E\left[e^{tX}\right] = \sum_{k=0}^{\infty} e^{tk} \cdot e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda e^t\right)^k}{k!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda\left(e^t-1\right)}.$$

Thus

$$P\left(X \geq a\right) \leq e^{\lambda\left(e^t-1\right)-ta},\ \forall t > 0.$$

When $t = \ln\frac{a}{\lambda}$, we can get $\min_{t>0} e^{\lambda\left(e^t-1\right)-ta} = \frac{e^{-\lambda}(e\lambda)^a}{a^a}$, thus

$$P(X \geq a) \leq \frac{e^{-\lambda}\left(e\lambda\right)^a}{a^a}.$$

(b) When $\sigma \geq a$, we have $\frac{2\sigma^2}{\sigma^2+a^2} \geq 1$, and it follows that

$$P\left(|X - E[X]| \geq a\right) \leq 1 \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$

When $\sigma < a$, using Markov's inequality, we have

$$P\left(|X - E[X]| \geq a\right) = P\left(\left(X - E[X]\right)^2 \geq a^2\right) \leq \frac{E\left[\left(X - E[X]\right)^2\right]}{a^2} = \frac{\sigma^2}{a^2}.$$

Since $\sigma < a$, we have $\frac{\sigma^2}{a^2} \leq \frac{2\sigma^2}{\sigma^2+a^2}$, thus

$$P\left(|X - E[X]| \geq a\right) \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$