Answer 1

Probaility of fair, six-sided dice. Let $A, B \in \{1, 2, ..., 6\}$ represent the events from the first and second coin flip respectively.

We want to find: $\mathsf{p}(A = 5 \text{ or } B = 5)$

We know that $\mathsf{p}(A \text{ or } B) = \mathsf{p}(A) + \mathsf{p}(B) - \mathsf{p}(A \text{ and } B)$

So,

$$\mathsf{p}(A = 5 \text{ or } B = 5) = \mathsf{p}(A = 5) + \mathsf{p}(B = 5) - \mathsf{p}(A = 5 \text{ and } B = 5) \tag{1}$$

$$\mathsf{p}(A = 5) = \mathsf{p}(B = 5) = \frac{1}{6} \tag{2}$$

$$\mathsf{p}(A = 5 \text{ and } B = 5) = \mathsf{p}(A = 5)\ \mathsf{p}(B = 5) = \frac{1}{36} \text{ (by independence)} \tag{3}$$

$$\text{Subsituting eqns (2) and (3) into (1)}$$

$$\Rightarrow \mathsf{p}(A = 5 \text{ or } B = 5) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$$

Answer 2

Bayes Rule. Let $X = \{0, 1\}$ represent if I have the disease or not where $X = 1$ means I have the disease and $X = 0$ means I do not have the disease. Also let $Y = \{\text{pos}, \text{neg}\}$ where $Y = \text{pos}$ means the test is positive and $Y = \text{neg}$ means the test is negative. We know:

$$\begin{aligned} \mathsf{p}(X = 1) &= 0.001 \\ \mathsf{p}(X = 0) &= 0.999 \\ \mathsf{p}(Y = \text{pos} \mid X = 1) &= 0.9 \\ \mathsf{p}(Y = \text{neg} \mid X = 1) &= 0.1 \\ \mathsf{p}(Y = \text{pos} \mid X = 0) &= 0.2 \\ \mathsf{p}(Y = \text{neg} \mid X = 0) &= 0.8 \end{aligned}$$

We want to find: $\mathsf{p}(X = 1 \mid Y = \text{pos})$. Using Bayes Rule:

$$\mathsf{p}(X = 1 \mid Y = \text{pos}) = \frac{\mathsf{p}(Y = \text{pos} \mid X = 1)\ \mathsf{p}(X = 1)}{\mathsf{p}(Y = \text{pos})} \tag{4}$$

Where,

$$\begin{aligned} \mathsf{p}(Y = \text{pos}) &= \mathsf{p}(Y = \text{pos} \mid X = 0)\ \mathsf{p}(X = 0) + \mathsf{p}(Y = \text{pos} \mid X = 1)\ \mathsf{p}(X = 1) \\ &= 0.2 \times 0.999 + 0.9 \times 0.001 \approx 0.2007 \end{aligned} \tag{5}$$

Now, eqn (4) is

$$\begin{aligned} \mathsf{p}(X = 1 \mid Y = \text{pos}) &= \frac{0.9 \times 0.001}{0.2007} \\ &\approx 0.0045 = 0.45\% \end{aligned}$$

Answer 3

Uniform Distribution. X is a collection of N i.i.d $\mathsf{Uniform}(a, b)$ random variables, where $a = 0$ and $b = \frac{1}{2}$.

(a) $X \sim \mathsf{Uniform}(x|a, b) = \frac{1}{b-a} = \frac{1}{0.5-0} = 2$
So, $\mathsf{p}(X = x|a, b) = 2 \quad \forall x \in (0, 0.5)$

(b) This is asking for the probablity of $X = 0.00027$ according to the pdf. So, $\mathsf{p}(X = 0.00027 \mid a, b) = 2$, since $0.00027 \in (0, 0.5)$

(c) This is asking for the probability of attaining a single value, which is zero since an integral over a single point evaluates to 0. $\mathsf{Pr}(X = 0.00027|a, b) = 0$.

Answer 4

Poisson Distribution. $X \sim \mathsf{Pois}(\lambda)$, where $\lambda > 0$.

$$\begin{aligned} \text{Exponential family reference:} &\quad h(x)\, \exp\{\eta\, T(x) - A(\eta)\} \qquad (6) \\ h(x) : \text{scaling constant} & \\ \eta : \text{natural parameter} & \\ T(x) : \text{sufficient statistics} & \\ A(\eta) : \text{log partition function} & \end{aligned}$$

(a) Exponential family form:

$$\begin{gathered} \exp\{x\, \ln(\lambda) - \lambda - \ln(x!)\} \\ \frac{1}{x!}\, \exp\{x\, \ln(\lambda) - \lambda\} \end{gathered}$$

(b) $T(x) \approx x$

(c) $A(\eta) = \exp\{\eta\}$

(d) Response function: $\lambda = \exp\{\eta\}$, also Link function: $\eta = \ln(\lambda)$

(e) $\mathbb{E}(X) = \frac{\partial A(\eta)}{\partial \eta} = \lambda$
$\mathsf{Var}(X) = \frac{\partial^2 A(\eta)}{\partial \eta^2} = \lambda$

(f)

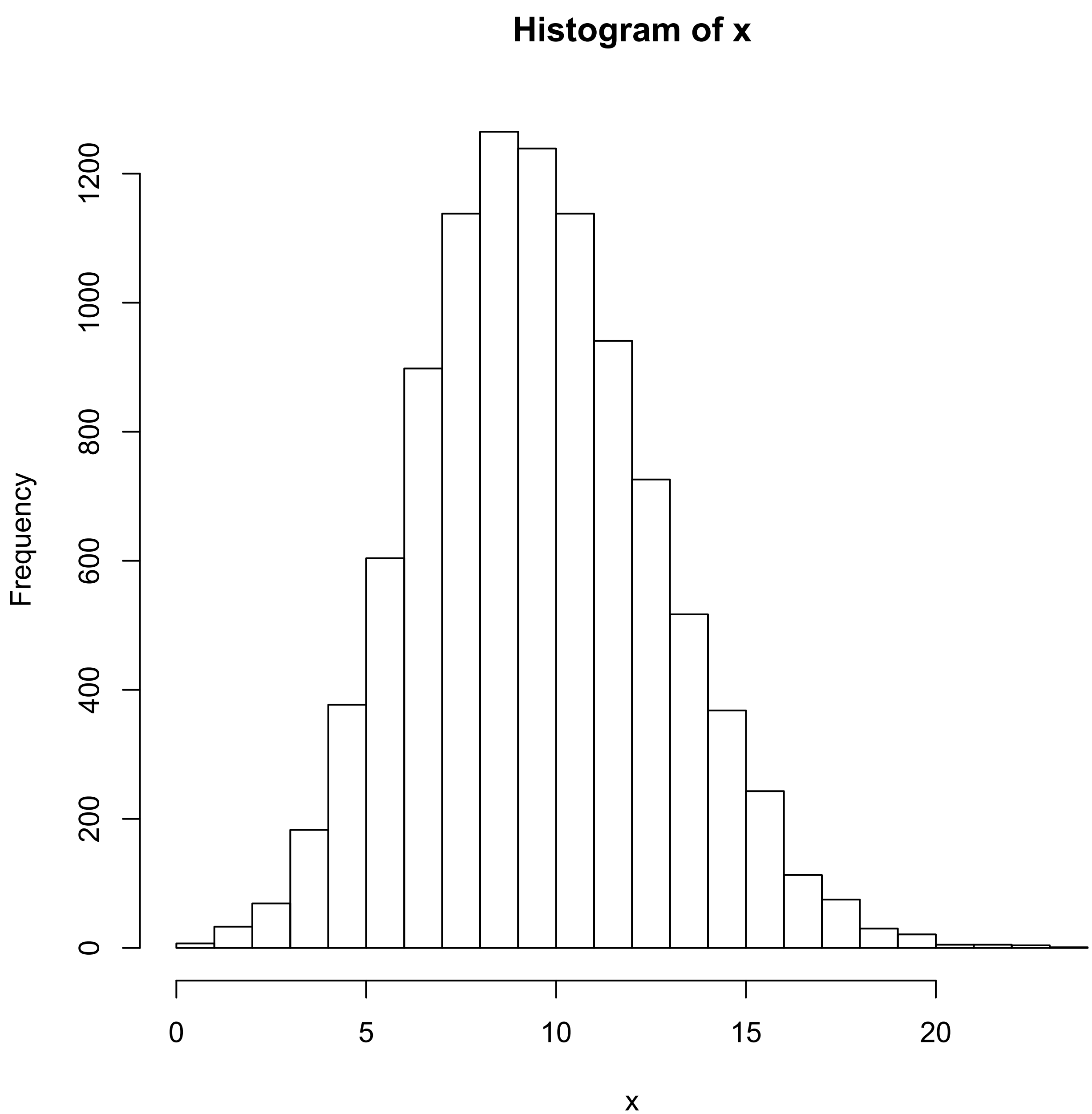
$$\begin{aligned}
\mathsf{p}(X \mid \lambda) &= \prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \\
\ln\, \mathsf{p}(X \mid \lambda) &= -n\lambda + \ln \lambda \sum_{i=1}^{n} X_i - \ln \prod_{i=1}^{n} X_i! \\
\frac{\partial\, \ln\, \mathsf{p}(X \mid \lambda)}{\partial \lambda} &= -n + \frac{\sum_{i=1}^{n} X_i}{\lambda} = 0 \\
\Rightarrow \hat{\lambda}_{MLE} &= \frac{\sum_{i=1}^{n} X_i}{n}
\end{aligned}$$

(g)

$$\begin{aligned}
\mathsf{p}(\lambda \mid X, \alpha, \beta) &= \mathsf{p}(X \mid \lambda, \alpha, \beta)\, \mathsf{p}(\lambda \mid \alpha, \beta) \\
&= \prod_{i=1}^{n} \frac{\lambda^{X_i}\, e^{-\lambda}}{X_i!}\; \frac{\lambda^{\alpha-1}}{e^{\beta\lambda}} \\
&= \left(\lambda^{\sum_{i=1}^{n} X_i + \alpha - 1}\, e^{-(\beta+n)\lambda} \right) / \prod_{i=1}^{n} X_i! \\
\ln\, \mathsf{p}(\lambda \mid X, \alpha, \beta) &= \left(\sum_{i=1}^{n} X_i + \alpha - 1 \right) \ln \lambda - (\beta + n)\lambda - \ln \prod_{i=1}^{n} X_i! \\
\frac{\partial\, \ln\, \mathsf{p}(\lambda \mid X, \alpha, \beta)}{\partial \lambda} &= \frac{\sum_{i=1}^{n} X_i + \alpha - 1}{\lambda} - (\beta + n) = 0 \\
\Rightarrow \hat{\lambda}_{MAP} &= \frac{\sum_{i=1}^{n} X_i + \alpha - 1}{\beta + n}
\end{aligned}$$

Answer 5

Extension to Problem 4, MAP and MLE simulation.



(a)

(b) 10.0171

(c) 10.0161

(d) 10.0260

(e) 10.0170

(f) We are trying to determine the best approximation of λ given the data X. Since we assumed our data is from the Poisson distribution, whose expectation is λ, and we know that the average of our data is 10.0171 (from part a), we would expect $\lambda \simeq 10.0171$. We assume the prior distribution of λ is the Gamma distribution, whose expectation given parameters α and β is $\frac{\alpha}{\beta}$. Therefore, the simulation that has $\alpha = 10$ and $\beta = 1$ should be a good approximation since $\frac{\alpha}{\beta}$ is the closest one to 10.0171. Also another thing to note is that we have a large sample size of our data ($n = 10,000$ time points), which according to our MAP estimate and initializations for both α and β, decreases the emphasis on our prior distribution. This is why we do not see a drastic difference in the MLE and MAP estimates of λ in parts b-e. In fact, the data X was created from a λ value of 10; so the "uniformed" prior estimate with $\alpha = 1$ and $\beta = 1$ achieved the best result.

Answer 6

Will post the code for this after the next assignment.

Answer 7

MAP Estimator for Logistic Regression.

Since $P(y = 1|x)$ is a Bernoulli random variable (say $y = \{-1, 1\}$) rewrite the following generalized linear model

$$\begin{aligned} f(x) = \beta^T x &= \log\left(\frac{P(y=1|x)}{P(y=-1|x)}\right) \\ &= \log\left(\frac{P(y=1|x)}{1-P(y=1|x)}\right) \end{aligned}$$

which implies

$$\begin{aligned} P(y=1|x) &= \frac{1}{1+\exp(\beta^T x)} \\ P(y=-1|x) &= \frac{1}{1+\exp(-\beta^T x)} \\ P(y=\pm 1|x) &= \frac{1}{1+\exp(y\,\beta^T x)}. \end{aligned}$$

So the likelihood is

$$\prod_{i=1}^{n} \frac{1}{1+\exp(y_i\,\beta^T x_i)}$$

if we use the same prior as in the ridge regression setting we obtain the following MAP estimator

$$\hat{\beta}_{\text{MAP}} = \arg\min_{\beta} \left[n^{-1} \sum_{i=1}^{n} \log(1 + \exp(-y_i\, \beta^T x_i)) + \lambda \|\beta\|^2 \right],$$