Probability and Statistics

Tutorial 12

Q1: (Discrete Bayes) A factory produces "trick" coins. Your prior belief is that a coin is either a "Type A" (with $p(H) = 0.25$) or a "Type B" (with $p(H) = 0.75$). You believe $P(\text{Type A}) = 0.5$ and $P(\text{Type B}) = 0.5$. You pick one coin, flip it 3 times, and observe 2 Heads and 1 Tail (HHT).

a. What is the posterior probability that the coin was "Type A"?

b. What is the Maximum a Posteriori (MAP) estimate for $p(H)$?

Solution:. Let $p_A = 0.25$ and $p_B = 0.75$. Let D be the data (2 Heads in 3 flips). The likelihood function is the Binomial PMF: $L(D \mid p) = \binom{3}{2} p^2 (1-p)^1 = 3p^2(1-p)$. We use Bayes' theorem, $P(p \mid D) \propto L(D \mid p) \times P(p)$.

First, calculate the likelihood for each type:

$$L(D \mid p_A) = 3(0.25)^2(1-0.25)^1 = 3(0.0625)(0.75) = 0.140625 = \frac{9}{64}$$
$$L(D \mid p_B) = 3(0.75)^2(1-0.75)^1 = 3(0.5625)(0.25) = 0.421875 = \frac{27}{64}$$

Now, find the (unnormalized) posterior:

$$P(p_A \mid D) \propto L(D \mid p_A)P(p_A) = \frac{9}{64} \times 0.5 = \frac{9}{128}$$
$$P(p_B \mid D) \propto L(D \mid p_B)P(p_B) = \frac{27}{64} \times 0.5 = \frac{27}{128}$$

To normalize, find the "evidence" $P(D) = \frac{9}{128} + \frac{27}{128} = \frac{36}{128} = \frac{9}{32}$.

a. The posterior probability for "Type A" is:

$$P(p_A \mid D) = \frac{9/128}{36/128} = \frac{9}{36} = \mathbf{0.25}$$

(And $P(p_B \mid D) = \frac{27/128}{36/128} = \frac{27}{36} = 0.75$)

b. The MAP estimate is the parameter that maximizes the posterior probability. Since $P(p_B \mid D) > P(p_A \mid D)$, the MAP estimate is $\hat{\mathbf{p}}_{\mathbf{MAP}} = \mathbf{0.75}$.

Q2: (Continuous Bayes) Let the prior for a parameter p be the uniform distribution $p \sim U[0,1]$. We observe a single data point $y = 3$ from a Geometric distribution, which has the likelihood function $P(y \mid p) = p(1-p)^{y-1}$. Find the full posterior density function $f(p \mid y = 3)$.

Solution:. We use Bayes' theorem for continuous parameters: $f(p \mid y) = \frac{P(y|p)f(p)}{P(y)}$.

1. **Prior:** $f(p) = 1$ for $p \in [0,1]$. 2. **Likelihood:** $P(y = 3 \mid p) = p(1-p)^{3-1} = p(1-p)^2$. 3. **Posterior (unnormalized):**

$$f(p \mid y = 3) \propto P(y = 3 \mid p)f(p) = p(1-p)^2 \times 1 = p(1-p)^2$$

We recognize this as the kernel of a Beta distribution, $Beta(\alpha, \beta)$, which has the form $p^{\alpha-1}(1-p)^{\beta-1}$. Here, $\alpha - 1 = 1 \implies \alpha = 2$ and $\beta - 1 = 2 \implies \beta = 3$. So, $f(p \mid y = 3) = C \cdot p(1-p)^2$, where C is the normalization constant.

4. **Evidence (Normalization):** $P(y = 3) = \int_0^1 P(y = 3 \mid p) f(p)\, dp = \int_0^1 p(1-p)^2\, dp$. This is the integral for the Beta function, $B(2,3) = \frac{\Gamma(2)\Gamma(3)}{\Gamma(2+3)} = \frac{1!\cdot 2!}{4!} = \frac{2}{24} = \frac{1}{12}$.

Alternatively, we can integrate directly:

$$\int_0^1 p(1-2p+p^2)\, dp = \int_0^1 (p-2p^2+p^3)\, dp = \left[\frac{p^2}{2} - \frac{2p^3}{3} + \frac{p^4}{4}\right]_0^1 = \frac{1}{2} - \frac{2}{3} + \frac{1}{4} = \frac{6-8+3}{12} = \frac{1}{12}$$

5. **Posterior (Full):**

$$f(p \mid y = 3) = \frac{p(1-p)^2}{1/12} = \mathbf{12p(1-p)^2}, \quad \text{for } p \in [0,1]$$

(This is the PDF for a $Beta(2,3)$ distribution).

Q3: (Gaussian-Gaussian Conjugate) A machine produces rods with length μ. The measurement error is $\sigma^2 = 1$, so a single measurement $x \sim N(\mu, \sigma^2 = 1)$. Your prior belief for the unknown mean μ is $N(\mu_0 = 10, \sigma_0^2 = 1)$. You take one measurement and get $x_1 = 12$. What is the posterior distribution for μ?

Solution:. We find the posterior by $f(\mu \mid x_1) \propto f(x_1 \mid \mu) \times f(\mu)$. We ignore the constant coefficients and focus on the exponents.

1. **Prior:** $f(\mu) \propto \exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right) = \exp\left(-\frac{(\mu-10)^2}{2\cdot 1}\right)$

2. **Likelihood:** $f(x_1 \mid \mu) \propto \exp\left(-\frac{(x_1-\mu)^2}{2\sigma^2}\right) = \exp\left(-\frac{(12-\mu)^2}{2\cdot 1}\right)$

3. **Posterior:** $f(\mu \mid x_1) \propto \exp\left(-\frac{(\mu-10)^2}{2}\right) \times \exp\left(-\frac{(12-\mu)^2}{2}\right)$

$$f(\mu \mid x_1) \propto \exp\left(-\frac{1}{2}\left[(\mu-10)^2 + (12-\mu)^2\right]\right)$$

Now, we expand the terms inside the exponent:

$$\begin{aligned}
(\mu^2 - 20\mu + 100) + (\mu^2 - 24\mu + 144) &= 2\mu^2 - 44\mu + 244 \\
&= 2(\mu^2 - 22\mu) + 244 \\
&= 2(\mu^2 - 22\mu + 121 - 121) + 244 \\
&= 2(\mu - 11)^2 - 242 + 244 \\
&= 2(\mu - 11)^2 + 2
\end{aligned}$$

We drop the constant term '2', so the exponent is proportional to $2(\mu - 11)^2$.

$$f(\mu \mid x_1) \propto \exp\left(-\frac{1}{2}\left[2(\mu-11)^2\right]\right) = \exp\left(-(\mu-11)^2\right)$$

We rewrite this in the standard Gaussian form $\exp\left(-\frac{(\mu-\mu_{\text{post}})^2}{2\sigma_{\text{post}}^2}\right)$.

$$\exp\left(-\frac{(\mu-11)^2}{1}\right) = \exp\left(-\frac{(\mu-11)^2}{2\cdot 0.5}\right)$$

By matching the terms, we find:

- $\mu_{\text{post}} = 11$
- $\sigma^2_{\text{post}} = 0.5$

The posterior distribution is $\mu \mid \mathbf{D} \sim \mathbf{N}(\mu = \mathbf{11}, \sigma^2 = \mathbf{0.5})$.

Q4: (Gamma-Poisson Conjugate) The number of cars arriving at a toll booth in an hour follows a Poisson(λ) distribution. Your prior belief for the unknown rate λ is $Gamma(\alpha = 2, \beta = 1)$. You observe the arrivals for $n = 3$ hours and count $x_1 = 2, x_2 = 3, x_3 = 4$ cars.

a. What is the posterior distribution for λ?
b. Find the Maximum a Posteriori (MAP) estimate $\hat{\lambda}_{MAP}$.
c. Find the Conditional Expectation (CE) estimate $\hat{\lambda}_{CE}$.

Solution:.

a. We find the posterior by $f(\lambda \mid D) \propto f(D \mid \lambda) \times f(\lambda)$.
1. **Prior:** The prior is $\lambda \sim Gamma(2, 1)$. The PDF is:

$$f(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda} = \lambda^{2-1} e^{-1\lambda} = \lambda^1 e^{-\lambda}$$

2. **Likelihood:** The data is $n = 3$ observations, $x_1 = 2, x_2 = 3, x_3 = 4$. The likelihood for independent Poisson samples is:

$$\begin{aligned} f(D \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} &= \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \cdot \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \cdot \frac{\lambda^{x_3} e^{-\lambda}}{x_3!} \\ &\propto (\lambda^{x_1} e^{-\lambda}) \cdot (\lambda^{x_2} e^{-\lambda}) \cdot (\lambda^{x_3} e^{-\lambda}) \\ &= \lambda^{x_1+x_2+x_3} e^{-3\lambda} \end{aligned}$$

Let $S = \sum x_i = 2 + 3 + 4 = 9$. The likelihood is $f(D \mid \lambda) \propto \lambda^9 e^{-3\lambda}$.
3. **Posterior:** We multiply the prior and the likelihood.

$$\begin{aligned} f(\lambda \mid D) &\propto f(D \mid \lambda) \times f(\lambda) \\ &\propto \left[\lambda^9 e^{-3\lambda}\right] \times \left[\lambda^1 e^{-\lambda}\right] \\ &= \lambda^{9+1} e^{-(3+1)\lambda} \\ &= \lambda^{10} e^{-4\lambda} \end{aligned}$$

This resulting kernel, $\lambda^{10} e^{-4\lambda}$, has the form of a Gamma distribution, $\lambda^{\alpha_{\text{post}}-1} e^{-\beta_{\text{post}}\lambda}$.

- $\alpha_{\text{post}} - 1 = 10 \implies \alpha_{\text{post}} = 11$
- $\beta_{\text{post}} = 4$

The posterior distribution is $\lambda \mid \mathbf{D} \sim \mathbf{Gamma}(\mathbf{11}, \mathbf{4})$.

b. The **MAP estimate** is the **mode** of the posterior distribution. To find it, we maximize the posterior PDF. It is easier to maximize the log-posterior:

$$\ell(\lambda) = \log(f(\lambda \mid D)) \propto \log(\lambda^{10} e^{-4\lambda}) = 10 \log(\lambda) - 4\lambda$$

We take the derivative with respect to λ and set it to 0:

$$\frac{d\ell}{d\lambda} = \frac{10}{\lambda} - 4 = 0$$

$$\frac{10}{\lambda} = 4 \implies \lambda = \frac{10}{4}$$

The MAP estimate is $\hat{\lambda}_{\mathbf{MAP}} = \mathbf{2.5}$.

c. The **CE estimate** is the **mean** (expected value) of the posterior distribution. We must compute $E[\lambda \mid D]$ for our posterior $Gamma(\alpha = 11, \beta = 4)$. The full PDF is $f(\lambda \mid D) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} = \frac{4^{11}}{\Gamma(11)}\lambda^{10}e^{-4\lambda}$.

$$\begin{aligned}
\hat{\lambda}_{CE} = E[\lambda \mid D] &= \int_0^\infty \lambda \cdot f(\lambda \mid D)\, d\lambda \\
&= \int_0^\infty \lambda \cdot \left[\frac{4^{11}}{\Gamma(11)}\lambda^{10}e^{-4\lambda}\right] d\lambda \\
&= \frac{4^{11}}{\Gamma(11)} \int_0^\infty \lambda^{11}e^{-4\lambda}\, d\lambda
\end{aligned}$$

We solve the integral by recognizing it as the kernel of a $Gamma(12, 4)$ distribution. We know that $\int_0^\infty \frac{4^{12}}{\Gamma(12)}\lambda^{11}e^{-4\lambda}\, d\lambda = 1$ (since it's a PDF). Therefore, $\int_0^\infty \lambda^{11}e^{-4\lambda}\, d\lambda = \frac{\Gamma(12)}{4^{12}}$.

Now substitute this back:

$$\hat{\lambda}_{CE} = \frac{4^{11}}{\Gamma(11)} \cdot \left[\frac{\Gamma(12)}{4^{12}}\right]$$

Using the property $\Gamma(z+1) = z\Gamma(z)$, we have $\Gamma(12) = 11 \cdot \Gamma(11)$.

$$\hat{\lambda}_{CE} = \frac{4^{11}}{\Gamma(11)} \cdot \frac{11 \cdot \Gamma(11)}{4^{12}} = \frac{11}{4}$$

The CE estimate is $\hat{\lambda}_{\mathbf{CE}} = \mathbf{2.75}$.

Q5: Suppose $D = \{x_1, \ldots, x_n\}$ is a data set consisting of independent samples of a Bernoulli random variable with unknown parameter θ, i.e., $f(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$ for $x_i \in \{0, 1\}$. We are also given that $\theta \sim U[0, 1]$. Obtain an expression for the posterior distribution on θ. Using this, obtain $\hat{\theta}_{MAP}$ and the conditional expectation estimator $\hat{\theta}_{CE}$.

(Hint: $\int_0^1 \theta^m(1-\theta)^r d\theta = \frac{m!r!}{(m+r+1)!}$)

Solution:

$$\begin{aligned}
f(\theta|x_1, \ldots, x_n) &= \frac{f(x_1, \ldots, x_n|\theta)}{f(x_1, \ldots, x_n)} \\
&= \frac{f(x_1, \ldots, x_n|\theta)p(\theta)}{\int_0^1 f(x_1, \ldots, x_n|\theta)p(\theta)d\theta} \\
&= \frac{\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}}{\int_0^1 \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}d\theta}
\end{aligned}$$

Now using the fact that for integral values m and r

$$\int_0^1 \theta^m (1-\theta)^r d\theta = \frac{m!r!}{(m+r+1)!}$$

and letting $s = \sum_{i=1}^n x_i$, the expression for the posterior becomes:

$$f(\theta|x_1, \ldots, x_n) = \frac{(n+1)!\theta^s(1-\theta)^{n-s}}{s!(n-s)!}$$

Now, solving for $\hat{\theta}_{MAP}$:

$$\begin{aligned}\hat{\theta}_{MAP} &= \arg\max_\theta P(x_1, \ldots, x_n|\theta) \\ &= \arg\max_\theta \theta^s(1-\theta)^{n-s}\end{aligned}$$

Taking the derivative and setting to 0, we get:

$$\begin{aligned}s\theta^{s-1}(1-\theta)^{n-s} - (n-s)\theta^s(1-\theta)^{n-s-1} &= 0 \\ \theta^{s-1}(1-\theta)^{n-s-1}[s(1-\theta)-(n-s)\theta] &= 0\end{aligned}$$

Since θ cannot take the value of 0 or 1 for maximizing the posterior, the second part of the above expression must go to 0.

$$\begin{aligned}s(1-\theta)-(n-s)\theta &= 0 \\ s - s\theta - n\theta + s\theta &= 0 \\ \theta &= \frac{s}{n}\end{aligned}$$

We got that the MAP estimate for θ is $\frac{s}{n}$, which is just the sample mean $\bar{X}$ of the data. For $\hat{\theta}_{CE}$, we have:

$$\begin{aligned}\hat{\theta}_{CE} &= E[\theta|x_1, \ldots, x_n] \\ &= \int_0^1 \theta f(\theta|x_1, \ldots, x_n) d\theta \\ &= \frac{(n+1)!}{s!(n-s)!}\int_0^1 \theta^{s+1}(1-\theta)^{n-s}d\theta \\ &= \frac{(n+1)!}{s!(n-s)!}\frac{(s+1)!(n-s)!}{(n+2)!} \\ &= \frac{s+1}{n+2}\end{aligned}$$

Q6: Let $X \sim Uniform(0,1)$. Suppose that we know

$$Y|X = x \sim Geometric(x).$$

Find the posterior density of X given $Y = 2$, $f_{X|Y}(x|2)$.

Solution:

Using Bayes' rule we have

$$f_{X|Y}(x|2) = \frac{P_{Y|X}(2|x)f_X(x)}{P_Y(2)}.$$

We know $Y|X = x \sim Geometric(x)$, so

$$P_{Y|X}(y|x) = x(1-x)^{y-1}, \quad \text{for } y = 1, 2, \ldots$$

Therefore,

$$P_{Y|X}(2|x) = x(1-x).$$

To find $P_Y(2)$, we can use the law of total probability

$$\begin{aligned}
P_Y(2) &= \int_{-\infty}^{\infty} P_{Y|X}(2|x)f_X(x)dx \\
&= \int_0^1 x(1-x) \cdot 1dx \\
&= \frac{1}{6}.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
f_{X|Y}(x|2) &= \frac{x(1-x)\cdot 1}{\frac{1}{6}} \\
&= 6x(1-x), \quad \text{for } 0 \le x \le 1.
\end{aligned}$$

Q7: Let X be a continuous random variable with the following PDF:

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Also, suppose that $Y|X = x \sim Geometric(x)$. Find the MAP estimate of X given $Y = 5$.

Solution: From Bayes' rule, we know that the posterior density for $0 \le x \le 1$ is:

$$f_{X|Y}(x|5) \propto P_{Y|X}(5|x)f_X(x) = 3x^2 \cdot x(1-x)^4 = 3x^3(1-x)^4$$

(and 0 otherwise), where the symbol $\propto$ means proportional to as a function of x. Therefore, the MAP estimate is given by

$$\hat{x}_{MAP} = \arg\max_x \{3x^3(1-x)^4\}$$

which can be found by setting the derivative of the argument equal to zero and solving for x:

$$\begin{aligned}
0 &= 9\hat{x}_{MAP}^2(1-\hat{x}_{MAP})^4 - 12\hat{x}_{MAP}^3(1-\hat{x}_{MAP})^3 \\
\Rightarrow \hat{x}_{MAP} &= \frac{3}{7}
\end{aligned}$$

This value is indeed in the interval $[0, 1]$.