3.7

[Wei] 3.7. We calculate two moments of X to give a simple expression of p.

$$\begin{aligned}\alpha_1 = EX &= \sum_{k=1}^{\infty} kP(X=k)\\ &= \sum_{k=1}^{\infty} -\frac{1}{\ln(1-p)}p^k\\ &= -\frac{p}{(1-p)\ln(1-p)},\\ \alpha_2 = EX^2 &= \sum_{k=1}^{\infty} k^2 P(X=k)\\ &= \sum_{k=1}^{\infty} -\frac{1}{\ln(1-p)}kp^k\\ &= -\frac{p}{(1-p)^2\ln(1-p)}.\end{aligned}$$

It is easy to observe that $\frac{\alpha_1}{\alpha_2} = 1-p$, or equivalently,

$$p = 1 - \frac{\alpha_1}{\alpha_2}.$$

Thus we can derive an MoM of p as

$$\hat{p}_{MoM} = 1 - \frac{\hat{\alpha}_1}{\hat{\alpha}_2} = 1 - \frac{\sum_i X_i}{\sum_i X_i^2}.$$

3.8

[Wei] 3.8. (1) $\hat{\sigma}^{(1)}_{MoM} = \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_i |X_i|$.

(2) $\hat{\sigma}^{(2)}_{MoM} = \sqrt{\frac{1}{n}\sum_i (X_i - \bar{X})^2}$, where $\bar{X} = \frac{1}{n}\sum_i X_i$.

3.9

[Wei] 3.9. Since $EX_1 = a$ and $Var(X_1) = \sigma^2$, we have MoMs of a and σ that

$$\begin{cases} \hat{a}_{MoM} = \bar{X}, \\ \hat{\sigma}_{MoM} = \sqrt{S_n}, \end{cases}$$

where $\bar{X} = \frac{1}{n}\sum_i X_i$, $S_n = \frac{1}{n}\sum_i (X_i - \bar{X})^2$.

Notice that $P(X>1) = P\left(\frac{X-a}{\sigma} > \frac{1-a}{\sigma}\right) = \Phi\left(\frac{a-1}{\sigma}\right)$, we derive an MoM of $P(X>1)$

$$\widehat{P(X>1)}_{MoM} = \Phi\left(\frac{\hat{a}_{MoM} - 1}{\hat{\sigma}_{MoM}}\right) = \Phi\left(\frac{\bar{X}-1}{\sqrt{S_n}}\right).$$

3.13

[Wei] 3.13. (1) 总体的一二阶矩为

$$\begin{aligned}\alpha_1 = EX &= \int_{-\infty}^{+\infty} \frac{t}{2\sigma} e^{-\frac{|t-a|}{\sigma}} dt \\ &= \int_{-\infty}^{+\infty} \left(\frac{t-a}{2\sigma} + \frac{a}{2\sigma}\right) e^{-\frac{|t-a|}{\sigma}} dt \\ &= a, \\ \alpha_2 = EX^2 &= \int_{-\infty}^{+\infty} \frac{t^2}{2\sigma} e^{-\frac{|t-a|}{\sigma}} dt \\ &= \int_{-\infty}^{+\infty} \frac{(t-a+a)^2}{2\sigma} e^{-\frac{|t-a|}{\sigma}} dt \\ &= \int_{-\infty}^{+\infty} \left(\frac{(t-a)^2}{2\sigma} + \frac{a^2}{2\sigma}\right) e^{-\frac{|t-a|}{\sigma}} dt \\ &= a^2 + 2\sigma^2.\end{aligned}$$

那么

$$\begin{cases} a = \alpha_1, \\ \sigma = \sqrt{\frac{1}{2}(\alpha_2 - \alpha_1^2)} = \sqrt{\frac{1}{2}\mu_2}. \end{cases}$$

因此 a 和 σ 的矩估计为

$$\begin{cases} \hat{a}_{MoM} = \bar{X}, \\ \hat{\sigma}_{MoM} = \sqrt{\frac{1}{2}S_n}, \end{cases}$$

这里 $\bar{X} = \frac{1}{n}\sum_i X_i$, $S_n = \frac{1}{n}\sum_i (X_i - \bar{X})^2$.

(2) 对数似然为

$$l(a,\sigma) = -n\ln 2 - n\ln\sigma - \frac{\sum_i |x_i - a|}{\sigma}.$$

任意固定 σ, 最大化 $l(a) = l(a,\sigma)$ 等价于最小化 $\sum_{i=1}^n |x_i - a|$, 则 $\hat{a} = m_n$, m_n 代表样本中位数，且 $n = 2k+1$ 时 $m_n = X_{(k+1)}$, $n = 2k$ 时 $m_n \in [X_{(k)}, X_{(k+1)}]$

(注：利用中位数的性质:以 F_n 表示样本分布，m_n 表示样本中位数，则 $F_n(m) \geq 1/2,\quad 1 - F_n(m-) \geq 1/2$. 容易验证 $m_n = argmin_a \sum_{i=1}^n |x_i - a|$)

令 $\frac{\partial l}{\partial \sigma} = 0$, 则 $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^n |X_i - \hat{a}| = \frac{1}{n}\sum_{i=1}^n |X_i - m_n|$.

此时 $\frac{\partial^2 l}{\partial \sigma^2}|_{\hat{\sigma}} = -\frac{n}{\hat{\sigma}^2} < 0$.

若 $X_1, \dots, X_n$ 不全相等，则 $\hat{\sigma} > 0$, 在参数空间内。否则考虑 $\vec{X}_{2k-1} = (1, \dots, 1 + 1/k, \dots, 1)$ (第 k 位取 1+1/k, 其余取 1), $\vec{X}_{2k} = (1, \dots, 1+1/k, 1+1/k, \dots, 1)$ (第 k 位和第 k+1 位取 1+1/k, 其余取 1) ,$k \geq 1$。则 $\sigma_{2k-1}, \sigma_{2k} > 0$, 且 $\lim_{k\to\infty} \hat{\sigma}_{2k-1} = \lim_{k\to\infty} \hat{\sigma}_{2k} = 0$

综上，极大似然估计为

$$\begin{cases} \hat{a}_{MLE} = m_n, \\ \hat{\sigma}_{MLE} = \frac{1}{n}\sum_{i=1}^n |X_i - m_n|, \end{cases}$$

m_n 代表样本中位数。

3.15

[Wei] 3.15. (1) 总体的一二阶矩为

$$\begin{aligned}
\alpha_1 = EX &= \int_\mu^{+\infty} \frac{t}{\sigma} e^{-\frac{t-\mu}{\sigma}} dt \\
&\overset{s:=\frac{t-\mu}{\sigma}}{=} \int_0^{+\infty} \left(s + \frac{\mu}{\sigma}\right) e^{-s} \sigma ds \\
&= \mu + \sigma, \\
\alpha_2 = EX^2 &= \int_\mu^{+\infty} \frac{t^2}{\sigma} e^{-\frac{t-\mu}{\sigma}} dt \\
&= \int_0^{+\infty} (\sigma s + \mu)^2 e^{-s} ds \\
&= 2\sigma^2 + 2\sigma\mu + \mu^2, \\
\mu_2 = Var(X) &= \alpha_2 - \alpha_1^2 = \sigma^2
\end{aligned}$$

那么

$$\begin{cases} \mu = \alpha_1 - \sigma, \\ \sigma = \sqrt{\mu_2}. \end{cases}$$

因此 μ 和 σ 的矩估计为

$$\begin{cases} \hat{\mu}_{MoM} = \bar{X} - \sqrt{S_n}, \\ \hat{\sigma}_{MoM} = \sqrt{S_n}, \end{cases}$$

这里 $\bar{X} = \frac{1}{n}\sum_i X_i$, $S_n = \frac{1}{n}\sum_i (X_i - \bar{X})^2$.

(2) 对数似然函数为

$$l(\theta) = -n \ln \sigma - \frac{\sum_i x_i - n\mu}{\sigma}, \quad x_{(1)} \geq \mu.$$

固定 σ, 最大化 $l(\mu)$, 由定义取 $\hat{\mu}_{MLE} = X_{(1)}$. 由 $\frac{\partial l}{\partial \sigma} = 0$, 得 $\hat{\sigma} = \bar{X} - \hat{\mu}_{MLE} = \bar{X} - X_{(1)}$. 且 $\frac{\partial^2 l}{\partial \sigma^2}|_{\hat{\sigma}} = -\frac{n}{\hat{\sigma}^2} < 0$. (若 X_i 全相等, 可采取与之前类似的讨论, 此处略过). 故 $\hat{\sigma}_{MLE} = \bar{X} - X_{(1)}$.

(3) 由于 $P(X_1 \geq t) = e^{-\frac{t-\mu}{\sigma}}$, 则其矩估计和极大似然估计为

$$\widehat{P(X_1 \geq t)}_{MoM} = e^{-\frac{t-\hat{\mu}_{MoM}}{\hat{\sigma}_{MoM}}}.$$

$$\widehat{P(X_1 \geq t)}_{MLE} = e^{-\frac{t-\hat{\mu}_{MLE}}{\hat{\sigma}_{MLE}}}.$$

3.17

[Wei] 3.17. In this problem, we write the likelihood function as

$$L(\theta|\mathbf{X}) = \prod_{i=1}^n \mathbf{1}_{(\theta-1/2,\theta+1/2)}(X_i) = \mathbf{1}_{(X_{(n)}-1/2, X_{(1)}+1/2)}(\theta)$$

So, for any $0 < \lambda < 1$,

$$\hat{\theta}^*(\mathbf{X}) = \lambda(X_{(n)} - \frac{1}{2}) + (1-\lambda)(X_{(1)} + \frac{1}{2})$$

is an MLE estimator of θ.

3.21

[Wei] 3.21. Suppose the ratio of black and white balls is $\theta \in [0, 1]$. The likelihood function is

$$lik(\theta) = \binom{n}{k} \left(\frac{\theta}{\theta + 1}\right)^{n-k} \left(\frac{1}{\theta + 1}\right)^{k} = \binom{n}{k} \frac{\theta^{n-k}}{(\theta + 1)^n}.$$

Thus the log-likelihood function is

$$l(\theta) = \ln \binom{n}{k} + (n - k) \ln \theta - n \ln(\theta + 1).$$

If $k \neq 0$ or n, from $\frac{\partial l}{\partial \theta} = 0$, we obtain $\hat{\theta} = \frac{n-k}{k}$. Because $\frac{\partial^2 l}{\partial \theta^2}|_{\hat{\theta}} = k^2(\frac{1}{n} - \frac{1}{n-k}) < 0$, we have $\hat{\theta}_{MLE} = \frac{n-k}{k}$. If $k = 0$ or n, observe that l reaches its maximum at $\theta = +\infty$ or 0 respectively. In summary, if we denote $\frac{n}{0} := +\infty$, we have that $\hat{\theta}_{MLE} = \frac{n-k}{k}$ for all k. □

7.3

[Wei] 7.3. In this problem, suppose that the samples $X_1, \ldots, X_8$ *i.i.d.* $\sim$ *Bernoulli*(θ), where $X_i = 1$ if the product is useless, else $X_i = 0$. Then the observation $X = \sum_i X_i \sim B(8, \theta)$ denotes the number of useless products in those 8 samples.

Since $\theta = 0.1$ or 0.2 alternatively, we only need to calculate one of the posterior probability. For example,

$$\begin{aligned}
\pi(\theta = 0.1|X = 2) &= \frac{P(X = 2|\theta = 0.1)\pi(0.1)}{\sum_{i=0.1,0.2} P(X = 2|\theta = i)\pi(i)} \\
&= \frac{\binom{8}{2}0.1^2 0.9^6 0.7}{\binom{8}{2}0.1^2 0.9^6 0.7 + \binom{8}{2}0.2^2 0.8^6 0.3} \\
&= 0.5418.
\end{aligned}$$

Therefore, we derive that

$$\begin{aligned}
\pi(\theta = 0.2|X = 2) &= 1 - \pi(\theta = 0.1|X = 2) \\
&= 0.4582.
\end{aligned}$$

7.4

[Wei] 7.4. In this problem, suppose the observation $X \sim P(\lambda)$ denotes the number of errors in a record.

Since $\lambda = 1.0$ or 1.5 alternatively, we only need to calculate one of the posterior probability. For example,

$$\begin{aligned}
\pi(\lambda = 1.0|X = 3) &= \frac{P(X = 3|\lambda = 1.0)\pi(1.0)}{\sum_{i=1.0,1.5} P(X = 3|\lambda = i)\pi(i)} \\
&= \frac{e^{-1.0}\frac{1.0^3}{3!}0.4}{e^{-1.0}\frac{1.0^3}{3!}0.4 + e^{-1.5}\frac{1.5^3}{3!}0.6} \\
&= 0.2457.
\end{aligned}$$

Therefore, we derive that

$$\begin{aligned}
\pi(\lambda = 1.5|X = 3) &= 1 - \pi(\lambda = 1.0|X = 3) \\
&= 0.7543.
\end{aligned}$$

7.5

[Wei] 7.5. (1) We first calculate the kernel

$$\begin{aligned}\pi(\theta|x) &\propto p(x|\theta)\pi(\theta)\\ &\propto \frac{1}{\theta^2} I_{(x,1)}(\theta) I_{(0,1)}(\theta)\\ &\propto \frac{1}{\theta^2} I_{(x,1)}(\theta),\end{aligned}$$

where we use the fact that $2x$ is constant with respect to θ for the second line and $x \in (0,1)$ for the third line. Thus we get $\pi(\theta|x) = c(x)\frac{1}{\theta^2} I_{(x,1)}(\theta)$.

From

$$\int_x^1 \frac{1}{\theta^2} = \frac{1}{x} - 1,$$

and the normalization condition $\int \pi(\theta|x) = 1$, we have the posterior distribution of θ

$$\pi(\theta|x) = \frac{x}{(1-x)\theta^2} I_{(x,1)}(\theta), \quad 0 < x < 1.$$

□

(2) Similarly, first calculate the kernel

$$\begin{aligned}\pi(\theta|x) &\propto p(x|\theta)\pi(\theta)\\ &\propto \frac{1}{\theta^2} I_{(x,1)}(\theta) 3\theta^2 I_{(0,1)}(\theta)\\ &\propto 3 I_{(x,1)}(\theta),\end{aligned}$$

where we use the fact that $x \in (0,1)$ for the last line. Thus we get $\pi(\theta|x) = c(x) 3 I_{(x,1)}(\theta)$.

From

$$\int_x^1 3 = 3(1-x),$$

and the normalization condition $\int \pi(\theta|x) = 1$, we obtain the posterior distribution

$$\pi(\theta|x) = \frac{1}{(1-x)} I_{(x,1)}(\theta), \quad 0 < x < 1.$$

7.11

[Wei] 7.11. We say $\theta \sim Pareto(\theta_0, \alpha)$ if the density function of θ is in the form in the problem. If the prior distribution is $\theta \sim Pareto(\theta_0, \alpha)$, then the kernel of the posterior distribution is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto f_{\mathbf{X}|\theta}(\mathbf{x})\pi(\theta)\\ &\propto \frac{1}{\theta^n} I(0 < x_{(1)} \leq x_{(n)} < \theta) \times \frac{1}{\theta^{\alpha+1}} I_{(\theta_0,+\infty)}(\theta)\\ &\propto \frac{1}{\theta^{n+\alpha+1}} I_{(\tilde{\theta}_0,+\infty)}(\theta),\end{aligned}$$

where $\tilde{\theta}_0 = \max\{\theta_0, x_{(n)}\}$. Notice that the kernel is the same as that of $Pareto(\tilde{\theta}_0, \alpha+n)$. We conclude that the conjugate prior distribution family of θ is Pareto distribution. □

7.14

（1）

$$p(x|\theta)=\theta(1-\theta)^2,\pi(\theta)=1$$

$$\pi(\theta|x)\propto p(x|\theta)\pi(\theta)\propto\theta(1-\theta)^2\sim Be(2,3)$$

故后验期望估计为

$$E(\theta|x)=\hat{\theta}_E=\frac{2}{2+3}=0.4$$

（2）

$$p(x|\theta)=\theta^3(1-\theta)^7,\pi(\theta)=1,$$

$$\pi(\theta|x)\propto p(x|\theta)\pi(\theta)\propto\theta^3(1-\theta)^7\sim Be(4,8)$$

故后验期望估计为

$$E(\theta|x)=\hat{\theta}_E=\frac{4}{4+8}=\frac{1}{3}$$