Tutorial Worksheet 2 - Parametric Estimation

Problem 1. (Sampling Distributions)

(a) Let $T \sim t_1$ (the t distribution with 1 degree of freedom). Explain why T has the same distribution as $\frac{X}{|Y|}$ where $X, Y \overset{IID}{\sim} \mathcal{N}(0,1)$, and hence why T also has the same distribution as $\frac{X}{Y}$.
[Hints: The distribution of $\frac{X}{Y}$ when $X, Y \overset{IID}{\sim} \mathcal{N}(0,1)$ is also called the Cauchy distribution (the t distribution with 1 degree of freedom). You may check that it has PDF:

$$f(x) = \frac{1}{\pi} . \frac{1}{1+x^2}.$$

You may use this result without proof in part (b).]

(b) t_1 is an example of an extremely "heavy-tailed" distribution: For $T \sim t_1$, show that $\mathbb{E}[|T|] = \infty$ and $\mathbb{E}\left[T^2\right] = \infty$. If $T_1, \ldots, T_n \overset{IID}{\sim} t_1$, explain why the Law of Large Numbers and the Central Limit Theorem do not apply to the sample mean $\frac{1}{n}\left(T_1 + \ldots + T_n\right)$.

(c) Let $U_n \sim \chi_n^2$. Show that $1/\sqrt{\frac{1}{n} U_n} \to 1$ in probability as $n \to \infty$. (Hint: Apply the Law of Large Numbers and the Continuous Mapping Theorem.)

[**Continuous Mapping Theorem:** If random variables $\{X_n\}_{n=1}^{\infty}$ converge in probability to $c \in \mathbb{R}$ (as $n \to \infty$), and $g : \mathbb{R} \to \mathbb{R}$ is continuous, then $\{g(X_n)\}_{n=1}^{\infty}$ converge in probability to $g(c)$.]

(d) Using Slutsky's lemma, show that if $T_n \sim t_n$ for each $n = 1, 2, 3, \ldots$, then as $n \to \infty$, $T_n \to Z$ in distribution where $Z \sim \mathcal{N}(0,1)$. (This formalizes the statement that "the t_n distribution approaches to the standard normal distribution as n gets large".)

[**Slutsky's lemma:** If sequences of random variables $\{X_n\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=1}^{\infty}$ satisfy $X_n \to c$ in probability for a constant $c \in \mathbb{R}$ and $Y_n \to Y$ in distribution for a random variable Y, then $X_n Y_n \to cY$ in distribution.]

(e) If ξ_p is the p-th quantile of $F_{m,n}$ distribution (F distribution with (m, n) degree of freedoms) and ξ'_p the p-th quantile of $F_{n,m}$. Show that, $\xi_p \xi'_{1-p} = 1$.

Problem 2. (Methods of Estimation)

(a) Suppose $X_1, \ldots, X_n \overset{\text{IID}}{\sim}$ Geometric(p), where Geometric(p) is the geometric distribution on the positive integers $\{1, 2, 3, \ldots\}$ defined by the probability mass function (PMF)

$$f(x \mid p) = p(1-p)^{x-1},$$

with a single parameter $p \in [0,1]$. Compute the method-of-moments estimate of p, as well as the MLE of p. For large n, what approximately is the sampling distribution of the MLE? (You may use, without proof, the fact that the Geometric(p) distribution has mean $1/p$.)

(b) Let $X_1, \ldots, X_n \overset{\text{IID}}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$. We showed in class that the MLEs for μ and σ^2 are given by

$$\hat{\mu} = \bar{X} \text{ and } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2.$$

(i) By computing the Fisher information matrix $I\left(\mu, \sigma^2\right)$, derive the approximate joint distribution of $\hat{\mu}$ and $\hat{\sigma}^2$ for large n. (Hint: Substitute $v = \sigma^2$ and treat v as the parameter rather than σ.)

(ii) Suppose it is known that $\mu = 0$. Compute the MLE $\tilde{\sigma}^2$ in the one-parameter sub-model $\mathcal{N}\left(0, \sigma^2\right)$. The Fisher information matrix in part (i) has off-diagonal entries equal to 0 when $\mu = 0$ and n is large. What does this tell you about the standard error of $\tilde{\sigma}^2$ as compared to that of $\hat{\sigma}^2$?

(c) Let $X_1, \ldots, X_n \stackrel{IID}{\sim} \text{Uniform}(0, \theta)$ for a single parameter $\theta > 0$ having PDF

$$f(x \mid \theta) = \frac{1}{\theta} \mathbb{1}\{0 \leq X \leq \theta\}.$$

(i) Compute the MLE $\hat{\theta}$ of θ. (Hint: Note that the PDFs $f(x \mid \theta)$ do not have the same support for all $\theta > 0$, and they are also not differentiable with respect to θ you will need to reason directly from the definition of MLE.)

(ii) If the true parameter is θ, explain why $\hat{\theta} \leq \theta$ always, and hence why it cannot be true that $\sqrt{n}(\hat{\theta} - \theta)$ converges in distribution to $\mathcal{N}(0, v)$ for any $v > 0$.

(d) Consider a parametric model $\{f(x \mid \theta) : \theta \in \mathbb{R}\}$ of the form

$$f(x \mid \theta) = e^{\theta T(x) - A(\theta)} h(x),$$

where T, A, and h are known functions.

(i) Show that the Poisson(λ) model is of this form, upon reparametrizing by $\theta = \log \lambda$. What are the functions $T(x), A(\theta)$, and $h(x)$?

(ii) For any model of the above form, differentiate the identity

$$1 = \int e^{\theta T(x) - A(\theta)} h(x) dx$$

with respect to θ on both sides, to obtain a formula for $\mathbb{E}_\theta[T(X)]$, where $\mathbb{E}_\theta$ denotes expectation when $X \sim f(x \mid \theta)$. Verify that this formula is correct for the Poisson example in part (i).

(iii) The generalized method-of-moments estimator is defined by the following procedure: For a fixed function $g(x)$, compute $\mathbb{E}_\theta[g(X)]$ in terms of θ, and take the estimate $\hat{\theta}$ to be the value of θ for which

$$\mathbb{E}_\theta[g(X)] = \frac{1}{n} \sum_{i=1}^{n} g\left(X_i\right).$$

The method-of-methods estimator discussed in class is the special case of this procedure for $g(x) = x$. Let $X_1, \ldots, X_n \stackrel{IID}{\sim} f(x \mid \theta)$, where $f(x \mid \theta)$ is of the given form, and consider the generalized method-of-moments estimator using the function $g(x) = T(x)$. Show that this estimator is the same as the MLE. (You may assume that the MLE is the unique solution to the equation $0 = l'(\theta)$, where $l(\theta)$ is the log-likelihood.)

(e) Suppose that X is a discrete random variable with

$$\begin{aligned}
\mathbb{P}[X = 0] &= \frac{2}{3}\theta \\
\mathbb{P}[X = 1] &= \frac{1}{3}\theta \\
\mathbb{P}[X = 2] &= \frac{2}{3}(1 - \theta) \\
\mathbb{P}[X = 3] &= \frac{1}{3}(1 - \theta)
\end{aligned}$$

where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution: $\{3, 0, 2, 1, 3, 2, 1, 0, 2, 1\}$. (For parts (i) and (ii), feel free to use any asymptotic approximations you wish, even though $n = 10$ here is rather small.)

(i) Find the method of moments estimate of θ, and compute an approximate standard error of your estimate using asymptotic theory.

(ii) Find the maximum likelihood estimate of θ and compute an approximate standard error of your estimate using asymptotic theory. (Hint: Your formula for the log-likelihood based on n observations $X_1, \ldots, X_n$ should depend on the numbers of 0's, 1's, 2's, and 3's in this sample.)

(f) Let $(X_1, \ldots, X_k) \sim \text{Multinomial}\,(n, (p_1, \ldots, p_k))$. (This is not quite the setting of n IID observations from a parametric model although you can think of $(X_1, \ldots, X_k)$ as a summary of n such observations $Y_1, \ldots, Y_n$ from the parametric model Multinomial $(1, (p_1, \ldots, p_k))$, where Y_i indicates which of k possible outcomes occurred for the i^{th} observation.) Find the MLE of p_i.

Problem 3. (Properties of Estimator)

(a) Suppose X_1, X_2, X_3 are independent normally distributed random variables with mean μ and variance σ^2. However, instead of X_1, X_2, X_3, we only observe $Y_1 = X_2 - X_1$ and $Y_2 = X_3 - X_2$. Which of the following statistics is sufficient for σ^2?
(i) $Y_1^2 + Y_2^2 - Y_1Y_2$ (ii) $Y_1^2 + Y_2^2 + 2Y_1Y_2$ (iii) $Y_1^2 + Y_2^2$ (iv) $Y_1^2 + Y_2^2 + Y_1Y_2$.

(b) Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter p; $0 \le p \le 1$. The bias of the estimator $\frac{\sqrt{n}+2\sum_{i=1}^n X_i}{2(n+\sqrt{n})}$ for estimating p is equal to
(i) $\frac{1}{\sqrt{n}+1}\left(p - \frac{1}{2}\right)$ (ii) $\frac{1}{n+\sqrt{n}}\left(\frac{1}{2} - p\right)$ (iii) $\frac{1}{\sqrt{n}+1}\left(\frac{1}{2} + \frac{p}{\sqrt{n}}\right) - p$ (iv) $\frac{1}{\sqrt{n}+1}\left(\frac{1}{2} - p\right)$.

(c) Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution with the probability density function;

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}}, & \text{if } x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 0$. Derive the Cramér-Rao lower bound for the variance of any unbiased estimator of θ. Hence, prove that $T = \frac{1}{n}\sum_{i=1}^n X_i$ is the uniformly minimum variance unbiased estimator of θ.

(d) Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N\left(\mu, \sigma^2\right)$ distribution, where both μ and σ^2 are unknown. Find the value of b that minimizes the mean squared error of the estimator $T_b = \frac{b}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}\right)^2$ for estimating σ^2, where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

(e) Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with the probability density function;

$$f(x;\lambda) = \begin{cases} \frac{2x}{\lambda}e^{-\frac{x^2}{\lambda}}, & \text{if } x > 0 \\ 0, & \text{otherwise,} \end{cases}$$

where $\lambda > 0$. Find the maximum likelihood estimator of λ and show that it is sufficient and an unbiased estimator of λ.

Problem 4. (Sufficiency Principle)

(a) Let X be a single observation from a population, belonging to the family $\{f_0(x), f_1(x)\}$, where

$$f_0(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \text{ and } f_1(x) = \frac{1}{\pi\left(1+x^2\right)}; \quad x \in \mathbb{R}.$$

Find a non-trivial sufficient statistic for the family of distribution.

(b) Let $X_1, X_2, \ldots, X_n$ be a random sample from the following pdf. Find the non-trivial sufficient statistic in each case: [Hints: If in the range of X_i, there is the parameter of the distribution present then we have to use the concept of Indicator function $\left(X_{(1)} \text{ or } X_{(n)}\right)$ or $\min_i \{X_i\}$ or $\max_i \{X_i\}$.]

(i) $f(x;\theta) = \begin{cases} \theta x^{\theta-1} & ; \text{ if } 0 < x < 1 \\ 0 & ; \text{ otherwise.} \end{cases}$

(ii) $f(x;\mu) = \frac{1}{|\mu|\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)}{2\mu^2}}; x \in \mathbb{R}$

(iii) $f(x;\alpha,\beta) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathbf{B}(\alpha,\beta)} & ; \text{ if } 0 < x < 1;\ \alpha,\ \beta\ >\ 0 \\ 0 & ; \text{ otherwise.} \end{cases}$

(iv) $f(x;\mu,\lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{(x-\mu)}{\lambda}} & ; \text{ if } x > \mu \\ 0 & ; \text{ otherwise.} \end{cases}$

(v) $f(x;\mu,\sigma) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} & ; \text{ if } x > 0 \\ 0 & ; \text{ otherwise.} \end{cases}$

(vi) $f(x;\alpha,\theta) = \begin{cases} \frac{\theta\alpha^\theta}{x^{\theta+1}} & ; \text{ if } x > \alpha \\ 0 & ; \text{ otherwise.} \end{cases}$

(vii) $f(x;\theta) = \begin{cases} \frac{2(\theta-x)}{\theta^2} & ; \text{ if } 0 < x < \theta \\ 0 & ; \text{ otherwise.} \end{cases}$

(c) If $f_\theta(x) = \frac{1}{2};\ \theta - 1 < x < \theta + 1$, then show that $X_{(1)}$ and $X_{(n)}$ are jointly sufficient for θ, $X_i \sim U(\theta-1, \theta+1)$.

(d) If a random sample of size $n \geqslant 2$ is drawn from a Cauchy distribution with PDF

$$f_\theta(x) = \frac{1}{\pi\left[1 + (x-\theta)^2\right]},$$

where $-\infty < \theta < \infty$, is considered. Then can you have a single sufficient statistic for θ?

Problem 5. (Uniformly Minimum Variance Unbiased Estimator)

(a) Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x;p) = \begin{cases} p(1-p)^x & ; x = 0, 1, \ldots \\ 0 & ; \text{ otherwise.} \end{cases}$

Show that unbiased estimator of p based on $T = \sum_{i=1}^n X_i$ is unique. Hence or otherwise find the UMVUE of p.

(b) Let X_1 and X_2 be two independent random variables having the same mean θ. Suppose that $E\left(X_1 - \theta\right)^2 = 1$ and $E\left(X_2 - \theta\right)^2 = 2$. For estimating θ, consider the estimators $T_\alpha\left(X_1, X_2\right) = \alpha X_1 + (1-\alpha)X_2, \alpha \in [0,1]$. The value of $\alpha \in [0,1]$, for which the variance of $T_\alpha\left(X_1, X_2\right)$ is minimum, equals

(i) $\frac{2}{3}$ (ii) $\frac{1}{2}$ (iii) $\frac{1}{4}$ (iv) $\frac{3}{4}$

(c) Let $X_1, X_2, \ldots, X_n$ be a random sample from

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)}, & \text{if } x > \theta \\ 0, & \text{otherwise.} \end{cases}$$

Show that $T = X_{(1)}$ is a complete sufficient statistic. Hence find the UMVUE of θ.

(d) Is the following families of distribution regular in the sense of Cramer & Rao? If so, find the lower bound for the variance of an unbiased estimator of θ based on a sample of size n. Also, find the UMVUE of θ for the PDF:

$$f(x,\theta) = \frac{e^{-\frac{x^2}{2\theta}}}{\sqrt{2\pi\theta}} \quad ; -\infty < x < \infty, \;\; \theta > 0.$$

(e) Based on a random sample $X_1, X_2, \ldots, X_n$ from Gamma(α). Obtain an estimator of $\psi_\alpha = \frac{\partial}{\partial\alpha} \ln \Gamma(\alpha)$ which attains CRLB and its variance.

Problem 6. (Finding Confidence Intervals)

(a) Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{U}(0,\theta)$, $\theta > 0$. Find a confidence interval for θ with confidence coefficient $(1-\alpha)$, based on $X_{(n)}$.

(b) Consider a random sample of size n from the rectangular distribution

$$f(x,\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{otherwise.} \end{cases}$$

If Y be the sample range then ξ is given by $\xi^{n-1}\left[n - (n-1)\xi\right] = \alpha$. Show that, Y and $Y\xi^{-1}$ are confidence limit to θ with confidence coefficient $(1-\alpha)$.

(c) Consider a random sample of size n from an exponential distribution, with PDF

$$f_X(x) = \begin{cases} \exp[-(x-\theta)] & \text{if } \theta < x < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Suggest a $100(1-\alpha)\%$ confidence interval for θ.

Problem 7. (A Heteroskedastic Linear Model)

Consider observed response variables $Y_1, \ldots, Y_n \in \mathbb{R}$ that depend linearly on a single covariate $x_1, \ldots, x_n$ as follows:

$$Y_i = \beta x_i + \varepsilon_i.$$

Here, the ε_i's are independent Gaussian noise variables, but we do not assume they have the same variance. Instead, they are distributed as $\varepsilon_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$ for possibly different variances $\sigma_1^2, \ldots, \sigma_n^2$. The unknown parameter of interest is β.

(a) Suppose that the error variances $\sigma_1^2, \ldots, \sigma_n^2$ are all known. Show that the MLE $\hat{\beta}$ for β, in this case, minimizes a certain weighted least-squares criterion, and derive an explicit formula for $\hat{\beta}$.

(b) Show that the estimate $\hat{\beta}$ in part (a) is unbiased, and derive a formula for the variance of $\hat{\beta}$ in terms of $\sigma_1^2, \ldots, \sigma_n^2$ and $x_1, \ldots, x_n$.

(c) Compute the Fisher information $I_{\mathbf{Y}}(\beta) = -\mathbb{E}_\beta\left[l''(\beta)\right]$ in this model (still assuming $\sigma_1^2, \ldots, \sigma_n^2$ are known constants). Show that the variance of $\hat{\beta}$ that you derived in part (b) is exactly equal to $I_{\mathbf{Y}}(\beta)^{-1}$.

In the remaining parts of this question, denote by $\tilde{\beta}$ the usual (unweighted) least-squares estimator for β, which minimizes $\sum_i \left(Y_i - \beta x_i\right)^2$. In practice, we might not know the values of $\sigma_1^2, \ldots, \sigma_n^2$, so we might still estimate β using $\tilde{\beta}$.

(d) Derive an explicit formula for $\tilde{\beta}$, and show that it is also an unbiased estimate of β.

(e) Derive a formula for the variance of $\tilde{\beta}$ in terms of $\sigma_1^2, \ldots, \sigma_n^2$ and $x_1, \ldots, x_n$. Show that when all error terms have the same variance σ_0^2, this coincides with the general formula $\sigma_0^2 \left(X^T X\right)^{-1}$ for the linear model.

(f) Using the Cauchy-Schwarz inequality $\left(\sum_i a_i^2\right)\left(\sum_i b_i^2\right) \geq \left(\sum_i a_i b_i\right)^2$ for any positive numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, compare your variance formulas from parts (b) and (e) and show directly that the variance of β is always at least the variance of $\hat{\beta}$. Explain, using the Cramer-Rao lower bound, why this is to be expected given your finding in part (c).

Problem 8. (The delta method for two samples)

Let $X_1, \ldots, X_n \overset{IID}{\sim}$ Bernoulli(p), and let $Y_1, \ldots, Y_m \overset{IID}{\sim}$ Bernoulli(q), where the X_i's and Y_i's are independent. For example, $X_1, \ldots, X_n$ may represent, among n individuals exposed to a certain risk factor for a disease, which individuals have this disease, and $Y_1, \ldots, Y_m$ may represent, among m individuals not exposed to this risk factor, which individuals have this disease. The odds-ratio

$$\frac{p}{1-p} \bigg/ \frac{q}{1-q}$$

provides a quantitative measure of the association between this risk factor and this disease. The log-odds-ratio is the (natural) logarithm of this quantity,

$$\log\left(\frac{p}{1-p} \bigg/ \frac{q}{1-q}\right).$$

(a) Suggest reasonable estimators $\hat{p}$ and $\hat{q}$ for p and q, and suggest a plugin estimator for the log-odds-ratio.

(b) Using the first-order Taylor expansion

$$g(\hat{p}, \hat{q}) \approx g(p, q) + (\hat{p} - p)\frac{\partial g}{\partial p}(p, q) + (\hat{q} - q)\frac{\partial g}{\partial q}(p, q)$$

as well as the Central Limit Theorem and independence of the X_i's and Y_i's, derive an asymptotic normal approximation to the sampling distribution of your plugin estimator in part (a).

(c) Give an approximate 95% confidence interval for the log-odds-ratio $\log \frac{p}{1-p} \Big/ \frac{q}{1-q}$. Translate this into an approximate 95% confidence interval for the odds-ratio $\frac{p}{1-p} \Big/ \frac{q}{1-q}$. (You may use a plugin estimate for the variance of the normal distribution that you derived in part (b).)

Problem 9. (Laplace distribution and Bayesian Analysis)

The double-exponential distribution with mean μ and scale b is a continuous distribution over $\mathbb{R}$ with PDF

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right).$$

It is sometimes used as an alternative to the normal distribution to model data with heavier tails, as this PDF decays exponentially in $|x - \mu|$ rather than in $(x - \mu)^2$.

(a) What are the MLEs $\hat{\mu}$ and $\hat{b}$ given data $X_1, \ldots, X_n$? Why is this MLE $\hat{\mu}$ more robust to outliers than the MLE $\hat{\mu}$ in the $\mathcal{N}\left(\mu, \sigma^2\right)$ model?

You may assume that n is odd and that the data values $X_1, \ldots, X_n$ are all distinct. (Hint: The log-likelihood is differentiable in b but not in μ. To find the MLE $\hat{\mu}$, you will need to reason directly from its definition.)

(b) Suppose it is known that $\mu = 0$. In a Bayesian analysis, let us model the scale parameter as a random variable B with prior distribution $B \sim \text{InverseGamma}(\alpha, \beta)$, where $\alpha, \beta > 0$. If $X_1, \ldots, X_n \sim \text{Laplace}\,(0, b)$ when $B = b$, what are the posterior distribution and posterior mean of B given the data $X_1, \ldots, X_n$?
(Hints: The InverseGamma (α, β) distribution is a continuous distribution on $(0, \infty)$ with PDF

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}$$

and with mean $\frac{\beta}{\alpha-1}$ when $\alpha > 1$.)

(c) Still supposing it is known that $\mu = 0$, what is the MLE $\hat{b}$ for b in this sub-model? How does this compare to the posterior mean from part (b) when n is large?

Problem 10. (Bayesian inference for Multinomial Proportions)

The Dirichlet $(\alpha_1, \ldots, \alpha_K)$ distribution with parameters $\alpha_1, \ldots, \alpha_K > 0$ is a continuous joint distribution over K random variables $(P_1, \ldots, P_K)$ such that $0 \leq P_i \leq 1$ for all $i = 1, \ldots, K$ and $P_1 + \ldots + P_K = 1$. It has (joint) PDF

$$f\left(p_1, \ldots, p_k \mid \alpha_1, \ldots, \alpha_K\right) \propto p_1^{\alpha_1 - 1} \times \ldots \times p_K^{\alpha_K - 1}.$$

Letting $\alpha_0 = \alpha_1 + \ldots + \alpha_K$, this distribution satisfies

$$\mathbb{E}\left[P_i\right] = \frac{\alpha_i}{\alpha_0}, \quad \operatorname{Var}\left[P_i\right] = \frac{\alpha_i\left(\alpha_0 - \alpha_i\right)}{\alpha_0^2\left(\alpha_0 + 1\right)}.$$

(a) Let $(X_1, \ldots, X_6) \sim \text{Multinomial}\left(n, (p_1, \ldots, p_6)\right)$ be the numbers of 1's through 6's obtained in n rolls of a (possibly biased) die. Let us model $(P_1, \ldots, P_6)$ as random variables with prior distribution Dirichlet $(\alpha_1, \ldots, \alpha_6)$. What is the posterior distribution of $(P_1, \ldots, P_6)$ given the observations $(X_1, \ldots, X_6)$? What is the posterior mean and variance of P_1?

(b) How might you choose the prior parameters $\alpha_1, \ldots, \alpha_6$ to represent a strong prior belief that the die is close to fair (meaning $p_1, \ldots, p_6$ are all close to $1/6$) ?

(c) How might you choose an improper Dirichlet prior to represent no prior information? How do the posterior mean estimates of $p_1, \ldots, p_6$ under this improper prior compare to the MLE?

Some useful distributional functions[1]:

1. If $Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$, then
$$Z = \frac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1)$$
2. $Y \sim \mathcal{N}(0,1) \Longrightarrow Y^2 \sim \chi^2(1)$
3. $Y \sim \mathcal{N}\left(\mu, \sigma^2\right) \Longrightarrow aY + b \sim \mathcal{N}\left(a\mu + b, a^2\sigma^2\right)$
4. $Y \sim \mathcal{U}(0,1) \Longrightarrow -\ln Y \sim \text{exponential}(1)$

   Generalization: $Y \sim \mathcal{U}(0,1) \Longrightarrow -\beta \ln Y \sim \text{exponential}(\beta)$

   Related: $Y \sim \text{beta}(\alpha, 1) \Longrightarrow -\ln Y \sim \text{exponential}(1/\alpha)$

   Related: $Y \sim \text{beta}(1, \beta) \Longrightarrow -\ln(1-Y) \sim \text{exponential}(1/\beta)$
5. $Y \sim \text{exponential}(\alpha) \Longrightarrow Y^{1/m} \sim \text{Weibull}(m, \alpha)$

   Related: $Y \sim \text{Weibull}(m, \alpha) \Longrightarrow Y^m \sim \text{exponential}(\alpha)$
6. $Y \sim \mathcal{N}\left(\mu, \sigma^2\right) \Longrightarrow e^Y \sim \text{lognormal}\left(\mu, \sigma^2\right)$ or equivalently if $U \sim \text{lognormal}\left(\mu, \sigma^2\right) \Longrightarrow \ln U \sim \mathcal{N}\left(\mu, \sigma^2\right)$
7. $Y \sim \text{beta}(\alpha, \beta) \Longrightarrow 1 - Y \sim \text{beta}(\beta, \alpha)$
8. $Y \sim \mathcal{U}(-\pi/2, \pi/2) \Longrightarrow \tan Y \sim \text{Cauchy}$
9. $Y \sim \text{gamma}(\alpha, \beta) \Longrightarrow cY \sim \text{gamma}(\alpha, \beta c)$, where $c > 0$

   Special case: $2Y/\beta \sim \chi^2(2\alpha)$
10. $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \text{Bernoulli }(p) \Longrightarrow \sum Y_i \sim b(n,p)$
11. $Y_i \sim \text{gamma}\left(\alpha_i, \beta\right), i = 1, 2, \ldots, n$ (mutually independent)
$$\Longrightarrow \sum Y_i \sim \text{gamma}\left(\sum \alpha_i, \beta\right)$$

    Special case: $\alpha_i = 1$, for $i = 1, 2, \ldots, n$. Then $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \text{exponential}(\beta) \Longrightarrow \sum Y_i \sim \text{gamma}(n, \beta)$

    Special case: $\alpha_i = \nu_i/2, \beta = 2$. Then $Y_i \sim \chi^2\left(\nu_i\right), i = 1, 2, \ldots, n$ (mutually independent) $\Longrightarrow \sum Y_i \sim \chi^2\left(\sum \nu_i\right)$

    Combination: If $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \text{exponential}(\beta)$, then
$$\frac{2\sum Y_i}{\beta} \sim \chi^2(2n)$$
12. $Y_i \sim \text{Poisson}\left(\lambda_i\right), i = 1, 2, \ldots, n$ (mutually independent)
$$\Longrightarrow \sum Y_i \sim \text{Poisson}\left(\sum \lambda_i\right)$$
13. $Y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right), i = 1, 2, \ldots, n$ (mutually independent)
$$\Longrightarrow \sum a_i Y_i \sim \mathcal{N}\left(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2\right)$$

    Special case: $\mu_i = \mu$ and $\sigma_i^2 = \sigma^2$, for $i = 1, 2, \ldots, n$. Then $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$
$$\Longrightarrow \sum a_i Y_i \sim \mathcal{N}\left(\mu \sum a_i, \sigma^2 \sum a_i^2\right)$$

[1] See also Univariate Distribution Relationships

Special case of iid result: If $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, then $\bar{Y} \sim \mathcal{N}\left(\mu, \sigma^2/n\right)$

Special case of iid result: If $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$, then $\sum Y_i \sim \mathcal{N}\left(n\mu, n\sigma^2\right)$

14. If $Y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right), i = 1, 2, \ldots, n$ (mutually independent), then

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i} \sim \mathcal{N}(0,1),$$

for $i = 1, 2, \ldots, n$. Therefore, $U = \sum Z_i^2 \sim \chi^2(n)$ because $Z_1^2, Z_2^2, \ldots, Z_n^2$ are iid $\chi^2(1)$

15. $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim}$ geometric$(p) \Longrightarrow U = \sum Y_i \sim \text{nib}(n, p)$

16. $Y_1, Y_2 \overset{iid}{\sim} \mathcal{N}(0,1) \Longrightarrow U = Y_1/Y_2 \sim$ Cauchy

17. $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim}$ exponential$(\beta) \Longrightarrow Y_{(1)} \sim$ exponential (β/n)

18. $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim}$ Weibull$(m, \alpha) \Longrightarrow Y_{(1)} \sim$ Weibull$(m, \alpha/n)$

19. If $Z \sim \mathcal{N}(0,1), W \sim \chi^2(\nu)$, and $Z \perp\!\!\!\perp W$, then

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t(\nu)$$

20. $Y_1, Y_2, \ldots, Y_n \sim$ iid $\mathcal{N}\left(\mu, \sigma^2\right)$

$$\Longrightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

21. If $W_1 \sim \chi^2\left(\nu_1\right), W_2 \sim \chi^2\left(\nu_2\right)$, and $W_1 \perp\!\!\!\perp W_2$, then

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F\left(\nu_1, \nu_2\right)$$

22. If $F \sim F\left(\nu_1, \nu_2\right)$, then $1/F \sim F\left(\nu_2, \nu_1\right)$

23. If $T \sim t(\nu)$, then $T^2 \sim F(1, \nu)$

24. If $W \sim F\left(\nu_1, \nu_2\right)$, then

$$\frac{\left(\nu_1/\nu_2\right) W}{1 + \left(\nu_1/\nu_2\right) W} \sim \text{beta}\left(\nu_1/2, \nu_2/2\right)$$