This is a take home midterm. You can use your notes, my online notes on canvas and the textbook. You are supposed to work on your own text without external help. I'll be available to answer question in person or via email. Please, write clealy and legibly and take a readable scan before uploading.

To solve the Exam problems, I have not collaborated with anyone nor sought external help and the material presented is the result of my own work.

Name (print): ________________________________

Question:	1	2	Total
Points:	60	50	110
Score:			

Question 1 . . . . *60 point*

Let X_i, $i = 1, \dots, N$ be a random sample where the X_i take only three possible values x_1, x_2 and x_3 with probability p_1, p_2 and p_3 respectively. Here $p_1 + p_2 + p_3 = 1$ and $\underline{p} = (p_1, p_2, p_3)$ is the vector of unknown parameters. Thus the p.m.f. of the X_i is

$$p(x|\underline{p}) = \begin{cases} p_k & x = x_k,\ k = 1, 2, 3 \\ 0 & \text{otherwise} \end{cases}$$

For notational simplicity, we will assume that $x_1 = 1$, $x_2 = 2$, $x_3 = 3$ but the discussion below is general.

(a) (10 points) Given $\underline{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, consider the family of distributions

$$g(\underline{p}; \underline{\alpha}) = \frac{1}{B(\underline{\alpha})} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} p_3^{\alpha_3 - 1}$$

defined on the set $p_1, p_2, p_3 > 0$ and $p_1 + p_2 + p_3 = 1$. The distribution g is called a Multivariate Beta distribution (MBD) or Dirichlet distribution.

Show that

$$B(\underline{\alpha}) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}.$$

(**Hint**: use the change of variables $y = p_2/(1 - p_1)$.)

Solution: We need to compute

$$B(\underline{\alpha}) = \int_{p_1 + p_2 < 1} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} (1 - p_1 - p_2)^{\alpha_3 - 1} dp_1 dp_2 = \int_0^1 p_1^{\alpha_1 - 1} \int_0^{1 - p_1} p_2^{\alpha_2 - 1} (1 - p_1 - p_2)^{\alpha_3 - 1} dp_2 dp_1$$

Calling $y_2 = p_2/(1 - p_1)$ and $y_1 = p_1$ we get

$$B(\underline{\alpha}) = \int_0^1 y_1^{\alpha_1 - 1} (1 - y_1)^{\alpha_1 + \alpha_2 - 1} \int_0^1 y_2^{\alpha_2 - 1} (1 - y_2)^{\alpha_3 - 1} dy_2 dy_1 = B(\alpha_1, \alpha_2 + \alpha_3) B(\alpha_2, \alpha_3).$$

From which the thesis follows using that

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

(b) (10 points) Show that the $g(\underline{p};\underline{\alpha})$ form a conjugate family of prior distribution for the sample $\mathbf{X}$. Assuming that the hyperparamter for the prior distribution are $\underline{\alpha}$, find the hyperparamter $\underline{\alpha}'$ for the posterior distribution.

Solution: Given a realization $\mathbf{x}$ of the sample, let $n_1(\mathbf{x})$ be the number of time $x_i = 1$, $n_2(\mathbf{x})$ the number of time $x_i = 2$, and $n_3(\mathbf{x})$ the number of time $x_i = 3$. Clearly $n_1(\mathbf{x}) + n_2(\mathbf{x}) + n_3(\mathbf{x}) = N$. We get

$$g(\underline{p}|\mathbf{x}) \propto p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_3^{\alpha_3-1} p_1^{n_1(\mathbf{x})} p_2^{n_2(\mathbf{x})} p_3^{n_3(\mathbf{x})}$$

so that we get

$$\underline{\alpha}'(\mathbf{x}) = (\alpha_1 + n_1(\mathbf{x}), \alpha_2 + n_2(\mathbf{x}), \alpha_3 + n_3(\mathbf{x}))$$

(c) (10 points) Find the Bayes estimator $\underline{\hat{p}}^B(\mathbf{X})$ for the quadratic error loss given by

$$L(\underline{p}, \underline{a}) = C_1(p_1 - a_1)^2 + C_2(p_2 - a_2)^2 + C_3(p_3 - a_3)^2]$$

where C_1, C_2, and C_3 are positive constants. (**Hint**: you can use the Lagrange multiplier method.)

Solution: We need to find a_k that minimize

$$\mathbb{E}(C_1(p_1 - a_1)^2 + C_2(p_2 - a_2)^2 + C_3(p_3 - a_3)^2 \mid \mathbf{X}) \tag{1}$$

under the condition $a_1 + a_2 + a_3 = 1$. We find the conditions

$$2C_k\mathbb{E}(p_k - a_k \mid \mathbf{X}) = \lambda$$

where λ is a Lagrange multiplier. This gives

$$a_k = \mathbb{E}(p_k \mid \mathbf{X}). \tag{2}$$

This is clearly a minimum since L is convex and positive.

Alternatively observe that the absolute minimum of (1), without any condition on the a_k, is attained when a_k satisfy (2). Since the a_k in (2) satisfy the condition $a_1 + a_2 + a_3 = 1$, then they represent also the minimum under such a condition.

We observe now that

$$\mathbb{E}_{\underline{\alpha}}(p_1) = \frac{1}{B(\alpha_1, \alpha_2, \alpha_3)} \int_{p_1+p_2<1} p_1 p_1^{\alpha_1-1} p_2^{\alpha_2-1}(1 - p_1 - p_2)^{\alpha_3-1} dp_1 dp_2 =$$
$$\frac{B(\alpha_1 + 1, \alpha_2, \alpha_3)}{B(\alpha_1 + 1, \alpha_2, \alpha_3)} = \frac{\Gamma(\alpha_1 + 1)}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1 + \alpha_2 + \alpha_3 + 1)} = \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3}$$

so that we get

$$\hat{p}_k^B(\mathbf{X}) = \frac{\alpha_k + n_k(\mathbf{X})}{N + \sum_k \alpha_k}.$$

(d) (10 points) Show that $\hat{\underline{p}}^B(\mathbf{X})$ found above is a consistent estimator, that is show that

$$\lim_{N\to\infty} \mathbb{P}(|\hat{p}_1^B(\mathbf{X}) - p_1| + |\hat{p}_2^B(\mathbf{X}) - p_2| + |\hat{p}_3^B(\mathbf{X}) - p_3| \geq \delta) = 0$$

for every $\delta > 0$.

Solution: From the LLN we have that

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\frac{n_k(\mathbf{X})}{N} - p_k\right| > \delta\right) = 0$$

which implies that

$$\lim_{N\to\infty} \mathbb{P}\left(|\hat{p}_k^B(\mathbf{X}) - p_k| > \delta\right) = 0\,.$$

Observe that

$$\mathbb{P}(|\hat{p}_1^B(\mathbf{X}) - p_1| + |\hat{p}_2^B(\mathbf{X}) - p_2| + |\hat{p}_3^B(\mathbf{X}) - p_3| \geq \delta) \leq \\ \mathbb{P}(|\hat{p}_1^B(\mathbf{X}) - p_1| > \delta/3) + \mathbb{P}(|\hat{p}_2^B(\mathbf{X}) - p_2| > \delta/3) + \mathbb{P}(|\hat{p}_3^B(\mathbf{X}) - p_3| > \delta/3)$$

from which the thesis follows immediately.

(e) (10 points) Find the MLE $\hat{\underline{p}}^L$ for $\underline{p}$ and compare it with $\hat{\underline{p}}_B$. Discuss the existence of improper priors for the Bayes estimator.

Solution:
We have

$$L(\underline{p}, \mathbf{x}) = p_1^{n_1(\mathbf{x})} p_2^{n_2(\mathbf{x})} p_3^{n_3(\mathbf{x})}$$

or

$$l(\underline{p}, \mathbf{x}) = \ln L(\underline{p}, \mathbf{x}) = n_1(\mathbf{x}) \ln p_1 + n_2(\mathbf{x}) \ln p_2 + n_3(\mathbf{x}) \ln p_3 \,.$$

Maximizing under the condition $p_1 + p_2 + p_3 = 1$ we get

$$\frac{n_k(\mathbf{x})}{p_k} = \lambda$$

or

$$\hat{p}_k^L(\mathbf{X}) = \frac{n_k(\mathbf{X})}{N}$$

Thus, like for the Beta distribution, the improper prior is $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

(f) (10 points) Compute the expected value and the variance of $\hat{p}_i^L$ and the covariance between $\hat{p}_i^L$ and $\hat{p}_j^L$. The CLT tell us that $\sqrt{N}(\hat{p}_1^L - p_1, \hat{p}_2^L - p_2)$ converges in distribution, as $N \to \infty$, to a pair of bivariate Normal r.v. (Z_1, Z_2). Write the joint p.d.f. of (Z_1, Z_2).

Solution: Let Y_i be the r.v. that is 1 if $X_i = 1$ and 0 if $X_i \neq 1$. We then have

$$n_1(\mathbf{X}) = \sum_{i=1}^{N} Y_i$$

and the Y_i are a Bernoulli random sample. It follows that

$$\mathbb{E}(\hat{p}_1^L) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_i) = p_1 \qquad \operatorname{var}(\hat{p}_1^L) = \frac{1}{N^2}\sum_{i=1}^{N}\operatorname{var}(Y_i) = \frac{p_1(1-p_1)}{N}.$$

and similarly

$$\mathbb{E}(\hat{p}_k^L) = p_k \qquad \operatorname{var}(\hat{p}_k^L)\frac{p_k(1-p_k)}{N}$$

Calling Z_i the r.v. that is 1 if $X_i = 2$ and 0 if $X_i \neq 2$ we get

$$\operatorname{cov}(\hat{p}_1^L, \hat{p}_2^L) = \frac{1}{N^2}\sum_{i=1}^{N}\operatorname{cov}(Y_i, Z_i) = -\frac{p_1 p_2}{N}$$

and thus

$$\operatorname{cov}(\hat{p}_k^L, \hat{p}_l^L) = -\frac{p_k p_l}{N}$$

We thus know that

$$\mathbb{E}(Z_i) = 0 \qquad \operatorname{var}(Z_i) = p_i(1-p_i) \qquad \operatorname{cov}(Z_1, Z_2) = -p_1 p_2$$

and setting

$$C = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 \\ -p_1p_2 & p_2(1-p_2) \end{pmatrix}$$

we get

$$f_{Z_1,Z_2}(z_1, z_2) = \frac{1}{2\pi\sqrt{\det C}} \exp\left(-\frac{1}{2}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}' C^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) =$$

$$\frac{1}{2\pi\sqrt{p_1p_2p_3(1-p_1)(1-p_2)}}$$

$$\exp\left(-\frac{1}{2p_3}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}' \begin{pmatrix} \frac{1}{p_1(1-p_1)} & \frac{1}{\sqrt{p_1p_2(1-p_1)(1-p_2)}} \\ \frac{1}{\sqrt{p_1p_2(1-p_1)(1-p_2)}} & \frac{1}{p_2(1-p_2)} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right)$$

Question 2 . . . . *50 point*

Let X_i, $i = 1, \dots, N$ be a random sample with distribution

$$f(x|\lambda) = \frac{\lambda^k}{k!} x^{k-1} e^{-\lambda x}$$

where $k \in \mathbb{N}$ is known while λ is to de determined.

(a) (10 points) Show that the Γ distributions form a conjugate family of prior for the the sample $\mathbf{X}$. If the prior distribution on λ is $\Gamma(\alpha, \beta)$ find α' and β' such that the posterior distribution is $\Gamma(\alpha', \beta')$

Solution: If the prior $\xi(\lambda)$ is $\Gamma(\alpha, \beta)$ we have

$$\xi(\lambda|\mathbf{x}) \propto \lambda^{\alpha-1} e^{\beta\lambda} \lambda^{kN} e^{-\lambda \sum_i x_i} = \lambda^{\alpha+kN-1} e^{-\lambda(\beta + \sum_i x_i)}$$

so that

$$\alpha' = \alpha + Nk \qquad \beta' = \beta + \sum_i x_i$$

(b) (10 points) Write an equation (in term of α and β) for the Bayes $\hat{\lambda}_B(\mathbf{X})$ estimator associated with the loss function:

$$L(\lambda, a) = (\lambda - a)^4$$

Solution: We need to minimize $\mathcal{L}_{\alpha,\beta}(a) = \mathbb{E}_{\alpha,\beta}\left((\lambda - a)^4\right)$ thus we need to solve

$$\mathcal{L}'_{\alpha,\beta}(a) = \mathbb{E}_{\alpha,\beta}\left((\lambda - a)^3\right) = a^3 - 3\frac{\alpha}{\beta}a^2 + 3\frac{\alpha(\alpha+1)}{\beta^2}a - \frac{\alpha(\alpha+1)(\alpha+2)}{\beta^3} = 0$$

(c) (10 points) Show that $\hat{\lambda}_B(\mathbf{X})$ is a consistent estimator.

Solution: Observe first that

$$\mathcal{L}''_{\alpha,\beta}(a) = 3\left(a - \frac{\alpha}{\beta}\right) + 3\frac{\alpha}{\beta^2} > 0$$

so that the equation $\mathcal{L}'_{\alpha,\beta}(a) = 0$ has a unique solution and $\hat{\lambda}^B(\mathbf{x})$ is well defined. Given a sample of size N, the posterior hyperparameters satisfy

$$\frac{\alpha'}{\beta'} = \frac{\alpha + Nk}{\beta + \sum_{i=1}^N X_i} \xrightarrow[N\to\infty]{p} \frac{k}{\mathbb{E}(X_i)} = \lambda\,.$$

Similarly we get

$$\frac{\alpha'(\alpha'+1)}{\beta'^2} \xrightarrow[N\to\infty]{p} \lambda^2 \qquad \frac{\alpha'(\alpha'+1)(\alpha'+2)}{\beta'^3} \xrightarrow[N\to\infty]{p} \lambda^3$$

Given δ let A_δ be the set $\mathbf{x}$ such that

$$\left|\frac{\alpha'}{\beta'} - \lambda\right| < \delta \qquad \left|\frac{\sqrt{\alpha'(\alpha'+1)}}{\beta'} - \lambda\right| < \delta$$
$$\left|\frac{\sqrt[3]{\alpha'(\alpha'+1)(\alpha'+2)}}{\beta'} - \lambda\right| < \delta\,.$$

If $\mathbf{x} \in A_\delta$, taking $a_+ = \lambda + C\sqrt[3]{\delta}$ we get

$$\begin{aligned}
&\mathcal{L}'_{\alpha',\beta'}(a_+) = a_+^3 - 3\frac{\alpha'}{\beta'}a_+^2 + 3\frac{\alpha'(\alpha'+1)}{\beta'^2}a_+ - \frac{\alpha'(\alpha'+1)(\alpha'+2)}{\beta'^3} \geq \\
&(\lambda + C\sqrt[3]{\delta})^3 - (\lambda + C\sqrt[3]{\delta})^2(\lambda+\delta)^2 + (\lambda + C\sqrt[3]{\delta})(\lambda-\delta)^2 - (\lambda+\delta)^3 = \\
&(C\sqrt[3]{\delta} + \delta)^3 - 2\lambda\delta(\lambda - C\sqrt[3]{\delta}) = C^3\delta - 2\lambda^2\delta + o(\delta)
\end{aligned}$$

so that taking $C^3 > \lambda^2$ and δ small enough we have $\mathcal{L}'_{\alpha',\beta'}(a_+) > 0$. Similarly for $\mathbf{x} \in A_\delta$ and $a_- = \lambda + C\sqrt[3]{\delta}$ we get $\mathcal{L}'_{\alpha',\beta'}(a_-) < 0$. Thus for every $\mathbf{x} \in A_\delta$ we have

$$|\hat{\lambda}^B(\mathbf{x}) - \lambda| \leq C\sqrt[3]{\delta}$$

and

$$\mathbb{P}(|\hat{\lambda}^B(\mathbf{x}) - \lambda| \leq \epsilon) \geq \mathbb{P}(A_{(\epsilon/C)^3}).$$

The thisis follow observing that for every η we have

$$\lim_{N\to\infty} \mathbb{P}(A_\eta) = 1$$

(d) (10 points) Find the ML estimator $\lambda_{ML}(\mathbf{X})$ for the random sample $\mathbf{X}$.

Solution: The likelihood function is given by:

$$L(\lambda, \mathbf{x}) = \frac{\lambda^{kN}}{(\Gamma(k))^N} \left(\prod_i x_i \right)^k e^{-\lambda \sum_i x_i}$$

or

$$l(\lambda, \mathbf{x}) = \ln L(\lambda, \mathbf{x}) = Nk \ln \lambda - \lambda \sum_i x_i - N \ln \Gamma(k) + k \sum_i \ln x_i$$

Differentiating we get

$$\lambda = \frac{k}{\overline{\mathbf{x}}}$$

as the unique critical point. This is clearly a maximum since $L(0, , \mathbf{x}) = L(\infty, , \mathbf{x}) = 0$ while $L(\lambda, \mathbf{x}) \geq 0$ for every λ. Thus we have

$$\hat{\lambda}_{ML}(\mathbf{X}) = \frac{k}{\overline{\overline{\mathbf{X}}}} \,.$$

(e) (10 points) Compute the Fisher information of X_i and use it to find the asymptotic distribution of $\lambda_{ML}(\mathbf{X})$.

Solution: Since

$$\frac{\partial^2}{\partial \lambda^2} \ln f(x|\lambda) = -\frac{k}{\lambda^2}$$

we get

$$I(\lambda) = \frac{k}{\lambda^2}\,.$$

Moreover we have

$$\hat{\lambda}_{ML}(\mathbf{X}) \simeq \mathcal{N}\left(\lambda, \frac{\lambda^2}{kN}\right).$$