MA40189 - Question Sheet Three

Simon Shaw, s.shaw@bath.ac.uk
https://moodle.bath.ac.uk/course/view.php?id=1179

2024/25 Semester II

Set: Problems class, Friday 28th February 2025.
Due in: Problems class, Friday 7th March 2025. Paper copies may be submitted in the problems class or directly to me either in lectures or my office, 4W4.10. PDF copies may be submitted to the portal available on the Moodle page.
Task: Attempt questions 1-2, 3(d). If you are feeling particularly worthy you may also consider 3(e); questions 3(a)-(c), and 4 are extra questions which may be discussed in the problems class.

1. Let $X_1, \ldots, X_n$ be conditionally independent given θ, so $f(x \,|\, \theta) = \prod_{i=1}^{n} f(x_i \,|\, \theta)$ where $x = (x_1, \ldots, x_n)$, with each $X_i \,|\, \theta \sim Po(\theta)$. Suppose we judge that $\theta \sim Gamma(\alpha, \beta)$.
   (a) Find the distribution of $\theta \,|\, x$.
   (b) Show that the posterior mean can be written as a weighted average of the prior mean, denoted θ_0, and the maximum likelihood estimate of θ, $\overline{x}$.
   (c) Let Z be a future (unobserved) observation. Find the mean and variance of the predictive distribution $Z \,|\, x$. You may assume that X and Z are conditionally independent given θ and that $Z \,|\, \theta \sim Po(\theta)$.
   (d) The data in the table below are the number of fatal accidents on scheduled airline flights between 1976 and 1985.

1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
24	25	31	31	22	21	26	20	16	22

   Let Z be the number of fatal accidents in 1986 and X_i the number of fatal accidents in 1975+i. Adopting the model given above, with $\alpha = 101$ and $\beta = 5$ and assuming that $Z \,|\, x$ may be approximated by a $N(E(Z \,|\, X), Var(Z \,|\, X))$, find an approximate 95% prediction interval for the number of accidents in 1986, that is an interval (z_L, z_U) such that $P(z_L < Z < z_U \,|\, x) = 0.95$.

2. Let $X_1, \ldots, X_n$ be a finitely exchangeable sequence of random variables and consider any $i \neq j \in \{1, \ldots, n\}$.
   (a) Explain why the marginal distribution of any X_i does not depend upon i so that $E(X_i)$ and $Var(X_i)$ do not depend upon i.
   (b) Explain why the joint distribution of any X_i, X_j does not depend upon either i or j so that $Cov(X_i, X_j)$ does not depend on either i or j.
   (c) Let $Y = \sum_{i=1}^{n} X_i$. By considering $Var(Y) \geq 0$, or otherwise, show that

$$Corr(X_i, X_j) \;=\; \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}} \;\geq\; \frac{-1}{n-1}.$$

3. Suppose $X \mid \mu \sim N(\mu, \sigma^2)$ and $Y \mid \mu, \delta \sim N(\mu + \delta, \sigma^2)$, where σ^2 is known and X and Y are conditionally independent given μ and δ.

   (a) Find the joint distribution of X and Y given μ and δ.

   (b) Consider the improper noninformative joint prior distribution

   $$f(\mu, \delta) \quad \propto \quad 1.$$

   Find, up to a constant of proportionality, the joint posterior distribution of μ and δ given x and y. Are $\mu \mid x, y$ and $\delta \mid x, y$ independent?

   (c) Find the marginal posterior distribution $f(\delta \mid x, y)$.

   (d) Find the marginal posterior distribution $f(\mu \mid x, y)$.

   (e) Consider a future observation Z where $Z \mid \mu, \delta \sim N(\mu - \delta, \sigma^2)$ and Z is conditionally independent of X and Y given μ and δ. Find the predictive distribution $f(z \mid x, y)$.
   [Hint: Use the distributions $f(z \mid \mu, \delta)$ and $f(\mu, \delta \mid x, y)$. You will need a double integral to integrate out both μ and δ.]

4. Let $X_1, \ldots, X_n$ be exchangeable so that the X_i are conditionally independent given a parameter θ. Suppose that each $X_i \mid \theta \sim U(0, \theta)$.

   (a) Show that $M = \max_i(X_i)$ is a sufficient statistic for $X = (X_1, \ldots, X_n)$.

   (b) Show that the Pareto distribution, $\theta \sim Pareto(a, b)$,

   $$f(\theta) \quad = \quad \frac{ab^a}{\theta^{a+1}}, \quad \theta \geq b$$

   is a conjugate prior distribution.