Name:

STAT 215 – Midterm 1

1. For four pairs of a rare specimen of bird that nested last season, the number of egges per nest n_i, $i = 1, \dots, 4$, and the number of eggs hatched y_i, $i = 1, \dots, 4$, are observed, providing the data $n_i = 2, 3, 3$ and 4 and $y_i = 1, 2, 3$ and 3. For a fifth pair nesting this season in similar conditions, $n_5 = 3$ eggs were observed and are about to hatch. Let θ be the probability that an egg is hatched.

1a. Obtain the likelihood function $p(y|\theta)$ for θ, based on the observations $\underline{\boldsymbol{y}} = (y_1, \dots, y_4)$;

Assume independent binomials:

$$p(\underline{\boldsymbol{y}}|\theta) = \prod_{i=1}^{4} Bin(y_i; n_i, \theta) = \prod_{i=1}^{4} C_{y_i}^{n_i} \cdot \theta^{\sum y_i} (1-\theta)^{\sum (n_i - y_i)}$$

1b. Assess a conjugate prior that in your opinon is adequate and calculate the posterior $p(\theta|\underline{\boldsymbol{y}})$.

Use a conjugate Beta prior $\theta \sim Be(a, b)$ with, for example, $a = 1, b = 1$ to get a uniform prior (other choices for a, b are fine).

$$p(\theta|\underline{\boldsymbol{y}}) \propto p(\theta)\; p(\underline{\boldsymbol{y}}|\theta) = Be(\underbrace{a + \sum_{i=1}^{4} y_i}_{a_1}, \underbrace{b + \sum_{i=1}^{4} (n_i - y_i)}_{b_1})$$

1c. Find the posterior predictive distribution $p(y_5|\underline{\boldsymbol{y}})$ for y_5, the number of eggs hatched in the 5-th nest.

$$p(y_5|\underline{\boldsymbol{y}}) = \int p(y_5|\theta)\; p(\theta|y) d\theta = \int Bin(y_5; n_5, \theta)\; Be(\theta; a_1, b_1) d\theta = BeBin(y_5; n_5, a_1, b_1)$$

$$= \frac{\Gamma(n_5 + 1)}{\Gamma(y_5 + 1)\Gamma(n_5 - y_5 + 1)} \frac{\Gamma(a_1 + y_5)\Gamma(n_5 - y_5 + b_1)}{\Gamma(a_1 + b_1 + n_5)}$$

See p483 for a discussion of the Beta-Binomial distribution.

2. Let $x_1, \ldots, x_n$ be a random sample of the $N(\theta, \sigma^2)$ distribution and consider the conjugate prior distribution for θ and σ^2:

$$p(\theta|\sigma^2) = N(\mu_0, 1\sigma^2/\kappa_0), \quad p(\sigma^2) = Inv\chi^2(\nu_0, \sigma_0^2)$$

2a. Find the (marginal) prior $p(\theta)$ (just writing down the correct distribution and stating the parameters is fine – no need to re-derive it).

$$p(\theta) = t_{\nu_0}(\mu_0, \sigma_0^2/\kappa_0).$$

(see, for example, p481).

2b. Determine the parameters $(\mu_0, \kappa_0, \nu_0, \sigma_0)$ of the prior distribution knowing that $E(\theta) = 0$, $Var(\theta) = 1.0$, $E(\sigma^2) = 2$ and $Var(\sigma^2) = 1$.

$E(\theta) = 0 \Rightarrow \mu_0 = 0.$

$$E(\sigma^2) = \frac{\nu_0}{\nu_0 - 2}\sigma_0^2 = 2 \text{ and } Var(\sigma^2) = \frac{2\nu_0^2}{(\nu_0-2)^2(\nu_0-4)}\sigma_0^4 = 1$$
$$\Rightarrow Var(\sigma^2)/[E(\sigma^2)]^2 = 1/4 \Rightarrow \nu_0 = 12, \sigma_0^2 = 5/3.$$

Finally, $Var(\theta) = \nu_0/(\nu_0 - 2)\ \sigma_0^2/\kappa_0 \Rightarrow \kappa_0 = 2$

2c. In a sample $\underline{\mathbf{x}} = (x_1, \ldots, x_n)$ of size $n = 10$, $\bar{x} = 1$ and $\sum_{i=1}^n (x_i - \bar{x})^2 = 8$, were observed. Determine the posterior distribution $p(\theta|\underline{\mathbf{x}})$ (state the name of the distribution and the values of all relevant parameters.

$p(\theta|\underline{\mathbf{x}}) = t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n)$ with
$\kappa_n = \kappa + n = 12$, $\nu_n = \nu_0 + n = 22$, $\mu_n = 5/6$ and $\sigma_n^2 = (28\frac{5}{3})/22 = 1.35.$

2d. Obtain $p(y|\underline{\mathbf{x}})$ where $y = x_{n+1}$ is a new observation taken from the same population.

$$p(y|\underline{\mathbf{x}}) = \int\int \underbrace{p(y|\theta,\sigma^2)\ p(\theta|\underline{\mathbf{x}},\sigma^2)}_{N(y;\theta,\sigma^2)N(\theta;n/(n+\kappa_0)\bar{x},\sigma^2/(n+\kappa_0))} d\theta \underbrace{p(\sigma^2|\underline{\mathbf{x}})}_{Inv\chi^2(\sigma^2;\nu_n,\sigma_n^2)} d\sigma^2 =$$
$$\int N[y; n/(n+\kappa_0)\bar{x}, \sigma^2(1 + 1/(n+\kappa_0))] Inv\chi^2(\sigma^2; \nu_n, \sigma_n^2)\ d\sigma^2 =$$
$$t_{\nu_n}[y; n\,\bar{x}/\kappa_n, \sigma_n^2(1 + 1/\kappa_n)]$$

3. Suppose that the lifetimes $x_1, \ldots, x_n$ of n bulbs are exponentially distributed with mean θ.

3a. Obtain Jeffrey's prior $p_J(\theta)$ for θ and sow that it is improper.

$d^2/d\theta^2 \log p(x_i|\theta) = -1/\theta^2$ and $J(\theta) = -E[d^2/d\theta^2 \; \log p(x_i|\theta)] = 1/\theta^2 \Rightarrow$
$p_J(\theta) = [J(\theta)]^{1/2} = 1/\theta$.
$p_J(\theta)$ is improper because $\int_0^\infty 1/\theta \; d\theta = \infty$.

3b. Let $y = \#\{x_i < t\}$ denote the number of failures at a fixed time t. Obtain the likelihood function $p(y|\theta)$.

$y \sim Bin(n, p)$ with $p = P(x_i < t) = 1 - \exp(-\theta t)$.

3c. Show that if no bulbs fail before a pre-specified time $t > 0$, then the posterior distribution is also improper.

$$p(\theta|y) \propto p(\theta)\, p(y|\theta) = n\frac{1}{\theta}\, e^{-n\theta t}$$

and

$$\int_0^\infty p(\theta|y) d\theta \geq c\, n \int_0^1 \frac{1}{\theta}\, e^{-nt} d\theta = \infty.$$

4. Let $\underline{\mathbf{x}}_n = (x_1, \dots, x_n)$ be a random sample from the $N(0, \theta^2)$ distribution.

4a. Show that the maximum likelihood estimator $\hat{\theta}_n$ is

$$\hat{\theta}_n^2 = \frac{1}{n} \sum x_i^2.$$

$$\log p(\underline{\mathbf{x}}_n|\theta) = -n \log \theta - \frac{1}{2\theta^2} \sum x_i^2$$

and

$$\frac{d}{d\theta} \log p(\underline{\mathbf{x}}_n|\theta) = -n/\theta + 2/(2\theta^3) \sum x_i^2 \Rightarrow \hat{\theta}_n^2 = \frac{1}{n} \sum x_i^2.$$

4b. Obtain the asymptotic posterior distribution of θ as $n \to \infty$.

Note: Since no prior is specified you need to use an asymptotic result which does not depend on the prior. The answer should depend only on the data $\underline{\mathbf{x}}_n$ (of course you can use notation like $\hat{\theta}_n^2$ if you define it).

Use, for example, the approximation of the posterior by a normal $N(\hat{\theta}_n, [I(\hat{\theta}_n)]^{-1})$, where $I(\hat{\theta}_n)$ is the expected Fisher information

$$I(\hat{\theta}_n) = -nE\left[\frac{d^2}{d\theta^2} \log p(x_i|\hat{\theta}_n)\right] = n\,2/\hat{\theta}_n^2 = 2n/\hat{\theta}_n^2$$

i.e. $p(\theta|\underline{\mathbf{x}}_n) \approx N[\hat{\theta}_n^2, \hat{\theta}_n^2/(2n)]$.

Name:

STAT 215 – Midterm 1
Problem 5

Take home problem. Tear this page off the exam, and submit by Thursday 5:05pm. You can either bring it to class and submit it after the lecture, or you can at any time before class drop it off in my mailbox (211 Old Chem).
No group work on this problem.

5. Suppose that the prior distribution for $(\theta_1, \dots, \theta_n)$ is such that the θ_i's constitute a random sample from a $N(\mu, \tau^2)$ distribution and $\mu|\tau^2 \sim N(\mu_0, \tau^2/c)$. Here, μ and τ are unknown parameters, c is a known constant.

Obtain the (marginal) prior distribution $p(\theta_1, \dots, \theta_n)$, and in particular, calculate the covariance between θ_i and θ_j, $1 \leq i, j \leq n$ with $i \neq j$, supposing that

5a. τ^2 is known.

5b. τ^2 is unknown with prior distribution $\tau^{-2} \sim Ga(\alpha, \beta)$.