UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Examination in: STK4011/9011 — Suggested solution.

Day of examination: Monday December 16th 2013.

Examination hours: 09.00 – 13.00.

This problem set consists of 5 pages.

Appendices: Selected definitions and theorems from Casella & Berger.
Table of common distributions from Casella & Berger.

Permitted aids: Approved calculator.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

a) We have

$$\begin{aligned} f(x|p) &= \mathrm{P}(X_i = x|p) = (1-p)^{x-1}p = \tfrac{p}{1-p}\exp(x\log(1-p)) \\ &= h(x)c(p)\exp(t(x)w(p)) \end{aligned}$$

with $h(x) = 1, c(p) = p/(1-p), w(p) = \log(1-p)$ and $t(x) = x$, thus the geometric distributions belongs to the exponential family of distributions.

The joint distribution of $\mathbf{X} = (X_1, \dots, X_n)$, can then be written as

$$f_n(\mathbf{x}|p) = \prod_{i=1}^{n} f(x_i|p) = (\frac{p}{1-p})^n \exp(\sum_{i=1}^{n} x_i \log(1-p)).$$

It follows by the factorization theorem that the statistic $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is sufficient. Furthermore it is complete for the parameter p since the parameter space $\Theta = \{p : 0 < p \leq 1\}$ contains an open set.

b) The log-likelihood for p based on $\mathbf{X}$ is given as

$$l(p) = \log(f_n(\mathbf{X}|p)) = n\log(\frac{p}{1-p}) + \log(1-p)\sum_{i=1}^{n} X_i$$

and its derivative becomes

$$l'(p) = \frac{n}{p(1-p)} - \frac{\sum_{i=1}^{n} X_i}{1-p}.$$

(Continued on page 2.)

Thus $l'(\hat{p}) = 0 \Leftrightarrow \hat{p} = \frac{n}{\sum_{i=1}^{n} X_i} = 1/\bar{X}$ is the MLE for p. (We should check that $\hat{p}$ is actually a maximum, but this follow automatically since the X_i are iid from an exponential family).

Since $\mu = \mathrm{E}[X_i] = 1/p$ we then directly get that its MLE is given by $\hat{\mu} = 1/\hat{p} = \bar{X}$.

Furthermore the MLE of $\sigma^2 = \mathrm{Var}(X_i) = \frac{1-p}{p^2}$ becomes $\hat{\sigma}^2 = \frac{1-\hat{p}}{\hat{p}^2} = \bar{X}^2(1 - 1/\bar{X})$.

Clearly $\hat{\mu} = \bar{X}$ is unbiased, $\mathrm{E}[\hat{\mu}] = \mathrm{E}[\bar{X}] = \mathrm{E}[X_i] = \mu$.

However, $\hat{p} = 1/\bar{X}$ and $\hat{\sigma}^2 = \bar{X}^2 - \bar{X}$ are biased. (For $\hat{p}$ we can use Jensen's inequality showing $\mathrm{E}[\hat{p}] > 1/\mathrm{E}\bar{X} = p$. For $\hat{\sigma}^2$ we get $\mathrm{E}[\hat{\sigma}^2] = \mathrm{E}[\bar{X}^2] - \mu = \mathrm{Var}[\bar{X}^2] + (\mathrm{E}[\bar{X}])^2 - \mu = \sigma^2/n + \mu^2 - \mu = \sigma^2(1 + 1/n) > \sigma^2$.)

c) The variance of the derivative of the log-likelihood becomes

$$\mathrm{Var}[l'(p)] = \frac{n\mathrm{Var}(X_i)}{(1-p)^2} = n\frac{\sigma^2}{(1-p)^2} = n\frac{1}{p^2(1-p)}$$

Thus the Cramér-Rao inequality says that for unbiased estimators $\mu^\star$ we have

$$\mathrm{Var}[\mu^\star] \geq \frac{(d\mu/dp)^2}{\mathrm{Var}[l'(p)]} = \frac{1}{p^4}\frac{p^2(1-p)}{n} = \frac{1-p}{np^2} = \frac{\sigma^2}{n}.$$

However, the unbiased $\hat{\mu} = \bar{X}$ has variance equal to this lower bound σ^2/n and is thus the uniformly minimum variance unbiased estimator (UMVUE) for μ.

d) The estimator of p given by

$$\tilde{p} = I(X_1 = 1) = \begin{cases} 1 & \text{if } X_1 = 1 \\ 0 & \text{otherwise} \end{cases}$$

has expectation $\mathrm{E}[\tilde{p}] = \mathrm{P}(X_1 = 1) = f(1|p) = p$ and so is $\tilde{p}$ is unbiased.

An improved unbiased estimator for p can now be constructed through the Rao-Blackwell technique

$$p^\star = \mathrm{E}[\tilde{p}|T(\mathbf{X})]$$

where $T = T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is the sufficient statistic for p from question (a). We get

$$\mathrm{P}(X_1 = 1|T = t) = \mathrm{P}[X_1 = 1|T(\mathbf{X}) = t] = \frac{\mathrm{P}(X_1 = 1, T = t)}{\mathrm{P}(T = t)}.$$

But $\mathrm{P}(T = t) = \binom{t-1}{n-1}(1-p)^{t-n}p^n$. It follows that

$$\begin{aligned} \mathrm{P}(X_1 = 1, T = t) &= \mathrm{P}(X_1 = 1, T - X_1 = t - 1) = \mathrm{P}(X_1 = 1)\mathrm{P}(T - X_1 = t - 1) \\ &= p\binom{t-2}{n-2}(1-p)^{t-1-(n-1)}p^{n-1} \end{aligned}$$

(Continued on page 3.)

since $T - X_1$ must have a negative binomial distribution with distribution $\mathrm{P}(T - X_1 = u) = \binom{u-1}{n-2}(1-p)^{u-(n-1)}p^{n-1}$ and is independent of X_1. Putting this together, we get

$$\mathrm{P}(X_1 = 1|T = t) = \frac{\binom{t-2}{n-2}(1-p)^{t-1-(n-1)}p^n}{\binom{t-1}{n-1}(1-p)^{t-n}p^n} = \frac{\binom{t-2}{n-2}}{\binom{t-1}{n-1}} = \frac{n-1}{t-1}$$

In other words we have $p^\star = \frac{n-1}{\sum_{i=1}^n X_i - 1} = (1 - 1/n)/(\bar{X} - 1/n)$.

This unbiased estimator is the uniform minimum variance unbiased estimator (UMVUE) for p since it is a function of the complete statistic $T = \sum_{i=1}^n X_i$.

e) In question (c) we found $\mathrm{Var}(l'(p)) = n\frac{1}{p^2(1-p)}$. By the large sample properties of MLE's it follows that the limit of the distribution of $\sqrt{n}(\hat{p} - p)$ is a normal distribution with expectation zero and variance $p^2(1-p)$.

   For (ii) $\sqrt{n}(\hat{\mu} - \mu)$ and (iii) $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ we can use a Delta-method argument giving that both quantities converge in distribution to normal distributions with expectations zero and variances given by respectively for (ii) $(d\mu/dp)^2 p^2(1-p) = (1-p)/p^2 = \sigma^2$ and for (iii) $(d\sigma^2/dp)^2 p^2(1-p) = (1-p)^3/p^4$

   This could alternatively be derived using the large sample distribution $\sqrt{n}(\tau(\hat{p}) - \tau(p))$ for MLE $\tau(\hat{p})$ for a parameter $\tau(p)$.

   These expressions all depend on only p which has consistent MLE $\hat{p} = 1/\bar{X}$. Inserting a consistent estimate of p into these large sample variances and dividing by n we get estimators of the large sample variances of $\hat{p}, \hat{\mu}$ and $\hat{\sigma}^2$.

f) The likelihood ratio statistic for a hypothesis

$$\mathrm{H}_0 : p = p_0 \text{ versus } \mathrm{H}_1 : p \neq p_0$$

   for a fixed value $0 < p_0 \leq 1$ is given as rejecting when the likelihood ratio

$$\lambda(\mathbf{x}) = \frac{L(p_0)}{L(\hat{p})}$$

   (with the likelihood $L(p) = \exp(l(p))$) is sufficiently small, say smaller than or equal a λ_0. And with level α we require that $\mathrm{P}(\lambda(\mathbf{x}) < \lambda_0 | p = p_0) \leq \alpha$

   In particular with n large we approximately have $-2\log(\lambda(\mathbf{x})) \sim \chi_1^2$. The rejection region is thus given as $R = \{\mathbf{x} : -2\log(\lambda(\mathbf{x})) > \gamma_\alpha$ where $\mathrm{P}(\chi_1^2 > \gamma_\alpha) = \alpha$.

   The corresponding asymptotic $(1-\alpha)$ confidence region for the parameter p is then given as the set of p_0 such that given $\mathbf{x}$ and

(Continued on page 4.)

letting the likelihood ratio statistic $-2\log(\lambda(\mathbf{x}))$ varying over p_0 we have $-2\log(\lambda(\mathbf{x})) \leq \gamma_\alpha$.

This follows by the equivalence between testing and confidence sets (In this specific case the set is an interval).

Problem 2

a) The joint distribution of X_1, X_2 and X_3 is given as

$$\begin{aligned} f(x_1, x_2, x_3) &= f(x_1|\lambda)f(x_2|2\lambda)f(x_3|3\lambda) = \tfrac{\lambda^{x_1}(2\lambda)^{x_2}(3\lambda)^{x_3}}{x_1!x_2!x_3!}\exp(-\lambda-2\lambda-3\lambda) \\ &= \tfrac{2^{x_2}3^{x_3}}{x_1!x_2!x_3!}\exp(-6\lambda)\exp((x_1+x_2+x_3)\log(\lambda)) \end{aligned}$$

which can be written on the factorization form $h(x_1, x_2, x_3)g(x_1+x_2+x_3|\lambda)$, thus $T = X_1 + X_2 + X_3$ is a sufficient statistic for λ.

The log-likelihood becomes

$$l(\lambda) = \log(f(X_1, X_2, X_3)) = K + T\log(\lambda) - 6\lambda$$

which is maximized for MLE $\hat{\lambda} = T/6$. We can note that $\hat{\lambda}$ is unbiased and has variance $\text{Var}(\hat{\lambda}) = \frac{\text{Var}(T)}{6^2} = \lambda/6$.

The Cramér-Rao inequality tells us that an unbiased estimator of λ has variance equal to or greater than one over the variance of the derivative of $l(\lambda)$. However, this derivative equals $l'(\lambda) = T/\lambda - 6$ and has variance $6/\lambda$. Thus $\hat{\lambda}$ has variance equal to the Cramér-Rao lower bound and no unbiased estimator can have a smaller variance. In conclusion $\hat{\lambda}$ is UMVUE.

b) Given observed values $\mathbf{x}$ of the random variable $\mathbf{X}$ the lambda has a posterior distribution

$$\pi(\lambda|\mathbf{x}) \propto f(\mathbf{x}|\lambda)\pi(\lambda|\alpha,\beta) \propto \lambda^{\alpha+t-1}\exp(-\lambda(\frac{1}{\beta}+6))$$

where $t = x_1 + x_2 + x_3$. This implies that $\lambda|\mathbf{x}$ follows a gamma-distribution with parameters $\alpha' = \alpha + t$ and $\beta' = 1/(6 + 1/\beta))$ and that

$$\begin{aligned} \tilde{\lambda} &= \text{E}[\lambda|\mathbf{x}] = \tfrac{(\alpha+t)}{6+1/\beta} \\ \tilde{\sigma}^2 &= \text{Var}[\lambda|\mathbf{x}] = \tfrac{(\alpha+t)}{(6+1/\beta)^2} \end{aligned}$$

When (i) $\alpha \to 0, \beta \to \infty$ and $\alpha\beta = \mu$, we see that $\tilde{\lambda} \to t/6 = \hat{\lambda}$, i.e. the MLE and $\tilde{\sigma}^2 \to t/6^2$

When (ii) $\alpha \to \infty, \beta \to 0$ and $\alpha\beta = \mu$ then $\tilde{\lambda} \to \alpha\beta = \mu$, i.e. the prior mean and $\tilde{\sigma}^2 \to 0$.

Thus under (i) the influence of the prior becomes infinitely small and the posterior mean converges to MLE with its variance estimate

(Continued on page 5.)

$\hat{\lambda}/6 = t/36$ and under (ii) the influence of the prior becomes infinitely large, thus the mean and variance of the posterior becomes the same as that of the prior.

Problem 3

Only STK9011

The incomplete data likelihood is given by

$$L_Y(\lambda) = \prod_{i=1}^{n} P(Y_i = 0|\lambda)^{1-Y_i} P(Y_i = 1|\lambda)^{Y_i} = \exp(-\lambda \sum_{i=1}^{n} z_i * (1-Y_i)) \prod_{i=1}^{n} (1-\exp(-\lambda x_i))^{Y_i}$$

The complete data log-likelihood is given as $l_X(\lambda) = n\log(\lambda) - \lambda \sum_{i=1}^{n} X_i$. The EM-algorithm for obtaining the MLE of λ based on $\mathbf{Y}$ using the complete data log-likelihood $l_X(\lambda)$ is given by the two steps

(i) E-step: Given current estimate λ' of λ and incomplete data $\mathbf{Y}$ find $\mathrm{E}[l_X(\lambda)|\lambda', \mathbf{Y}] = n\log(\lambda) - \lambda \sum_{i=1}^{n} \mathrm{E}[X_i|\lambda', \mathbf{Y}]$ and

(ii) M-step maximize $\mathrm{E}[l_X(\lambda)|\lambda', \mathbf{Y}]$ over λ to obtain updated estimate $\lambda'' = n/\sum_{i=1}^{n} \mathrm{E}[X_i|\lambda', \mathbf{Y}]$.

For the E-step note that when $Y_i = 0, X_i > z_i$

$$\mathrm{E}[X_i|\lambda', \mathbf{Y}] = \mathrm{E}[X_i|\lambda', Y_i = 0] = z_i + 1/\lambda'$$

since $X_i - \lambda'|X_i > \lambda' \sim \exp(\lambda')$. When $Y_i = 1, X_i < z_i$

$$\mathrm{E}[X_i|\lambda', \mathbf{Y}] = \mathrm{E}[X_i|\lambda', Y_i = 1] = \frac{\int_0^{z_i} x\lambda' \exp(-\lambda' x)dx}{1 - \exp(-\lambda' z_i)} = \frac{1 - \lambda' z_i \exp(-\lambda' z_i) - \exp(-\lambda' z_i)}{\lambda'(1 - \exp(-\lambda' z_i))}$$

END