Statistical and Mathematical Methods DS-GA 1002 November 12, 2015

Recitation Session 8 Solutions

1. *Estimating mean of Gaussian.*

   a. $Y = \mu + W$ where W is mean zero variance P.

   b. The ML estimator is the given value itself. By inspection of $f_Y(y|\mu) = \frac{1}{\sqrt{2}}\exp(-\frac{(y-\mu)^2}{2P})$, $\mu_{ML} = y$.

   c. Now the input is sum of two random variables $Y = M + W$. Given $Y = y$, the posterior is normal distribution with mean $\frac{Ny}{N+P}$ and variance $\frac{NP}{N+P}$. Mean and maximizing value of this density are equal, so here MMSE=MAP.

   d. Bayesian estimators are smaller than the ML estimator by a factor of $\frac{N}{N+P}$, if N is smaller compared to P, the prior information indicates that the μ is small. As N grows the prior distribution becomes flat, increasingly uniform, Bayesian estimator converges to the ML estimator.

2. *Estimating mean of Poisson.*

   a. Let Y be the signal. Given $Y = k$, the MLE is obtained by maximizing $p_Y(k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$ with respect to λ. Drop the denominator, take logarithm: $-\lambda + k\log\lambda$ therefore the MLE is k. By derivative test it is increasing in λ for $\lambda < k$, and decreasing beyond k.

   b. Maximize the posterior, $p_{\Lambda|Y}(\lambda|y)$, with respect to λ, it is proportional to:

   $$p_Y(k|\lambda)f_\Lambda(\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}\frac{I_{0<\lambda<\theta}}{\theta} \tag{1}$$

   The first factor is increasing in λ for $\lambda < k$. Therefore MAP is at $mink, \theta$.

   c. Prior gives us an upper bound for the parameter but no more than that. If θ is less than k then the MAP estimator is smaller than ML estimator. As θ goes to infinity MAP converges to ML.

   d. For an unknown but fixed value θ, say $\lambda < \theta$ then the ML estimator is naturally restricted, and gives the same formula as above for MAP.

3. *Functions of estimators.*

   a. Yes, because the transformation is linear, the pdf of $3 + 5\Theta$ is a scaled version of the pdf of Θ. Take squares for example, $f(x) = x^2$, it is not linear.

   b. Yes, because the MMSE estimator is given by the conditional expectation, which is linear. That is, $3 + 5E[\theta|Y] = \mathrm{E}[3 + 5\theta|Y]$. Again squares are not expected to give the same, if squares are the same then variance would be zero, which, in general, is not true.

   c. Yes, because the transformation is invertible. For example in question 1 part b maximizing $f_Y(y|g(\mu))$ will give $(g(\mu))_{ML} = y$. Note that taking a square if the argument is positive (or negative but not both) is invertible so the answer would be still yes. Take a degenerate transformation like $f(x) = 1$.