CSI 771 Problem Set 1

1.2 We have

$$\begin{aligned}
\mathrm{E}\Big(\big(P_n(y)\big)^2\Big) &= \frac{1}{n^2}\mathrm{E}\left(\left(\sum_{i=1}^n \mathrm{I}_{(-\infty,y]}(Y_i)\right)^2\right)\\
&= \frac{1}{n^2}\mathrm{E}\left(\sum_{i=1}^n \mathrm{I}_{(-\infty,y]}(Y_i) + \sum_{i\neq j}\mathrm{I}_{(-\infty,y]}(Y_i)\mathrm{I}_{(-\infty,y]}(Y_j)\right)\\
&= \frac{1}{n^2} n\mathrm{E}(\mathrm{I}_{(-\infty,y]}(Y_i)) + n(n-1)\mathrm{E}(\mathrm{I}_{(-\infty,y]}(Y_i))\mathrm{E}(\mathrm{I}_{(-\infty,y]}(Y_j))\\
&= \frac{1}{n}P(y) + \frac{n-1}{n}(P(y))^2.
\end{aligned}$$

Hence,

$$\mathrm{V}(P_n(y)) = \mathrm{E}\Big(\big(P_n(y)\big)^2\Big) - \Big(\mathrm{E}\big(P_n(y)\big)\Big)^2 = \frac{1}{n}P(y) + \frac{n-1}{n}(P(y))^2 - (P(y))^2 = \frac{P(y)(1-P(y))}{n}$$

1.3 We have

a.

$$\Psi(P) = \int\left(y - \int u\,\mathrm{d}P(u)\right)^2 \mathrm{d}P(y)$$

b.

$$\begin{aligned}
\Psi(P_n) &= \int_{-\infty}^{\infty}\left(y - \int_{-\infty}^{\infty} u\,\mathrm{d}P_n(u)\right)^2 \mathrm{d}P_n(y)\\
&= \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2.
\end{aligned}$$

c. $\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2$

d. It is biased (recall that $\frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2$ is unbiased), but the plug-in estimator is consistent. It is also the MLE for the variance of a normal random variable. (Notice that to form an MLE we need a likelihood, that is, a probability density function. Hence, being an "MLE" is rarely a property of just an estimator; it almost always requires specification of a distribution. Of course, this particular estimator is an MLE for the variance of many distributions other than the normal.)

1.7 **a.** Remember, "least squares" means to minimize the sum of squares in the differences in the fitted model values ("predictions") and the observed values. There are, of course, several things that the model predicts. The purpose of this exercise is to emphasize the underlying idea of least squares. The least squares estimators of the gamma parameters would very rarely be used.

A gamma model predicts that an observation will be $\alpha\beta$ and that the square of an observation will be $\alpha\beta^2 + (\alpha\beta)^2$. Hence we can form two different residuals $y_i - ab$ and $y_i^2 - ab^2 - (ab)^2$, and minimize the sums of their squares.

d. We maximize the log-likelihood function $l_L(\alpha, \beta|y_1, \ldots, y_n)$ w.r.t. α and β for a given sample, $y_1, \ldots, y_n$:

$$\max_{a,b}\left(-n\log\Gamma(a) - na\log(b) + (a-1)\sum\log(y_i) - \frac{1}{b}\sum y_i\right).$$

This does not have a closed-form solution. We can use Newton's method. We from the gradient by taking $\partial l_L/\partial a$, and $\partial l_L/\partial b$ and form the Hessian $\partial^2 l_L/\partial a\partial b$. The Hessian is negative definite. As the maximization problem is solved by Newton's method, the values of the gradient and Hessian are updated in each step. The inverse of the Hessian at the last step is an estimate of the variance-covariance of the estimators of α and β.

1.9a The MLE can be determined as the solution to

$$(-x_1 - (x_2 + x_3) - x_4)t^2 + (x_1 - 2(x_2 + x_3) - x_4)t + 2x_4 = 0.$$

The solution is $t = \hat{\theta} = 0.63$.