8.1 The Sampling Distribution

Previously, we've used statistics as means of estimating the value of a parameter, and have selected which statistics to use based on general principle: The *Bayes Estimator* was selected to minimize average squared loss, for a given prior distribution, while the *MLE* represented the value where the likelihood function attained a maximum, and the Method of Moments estimator was obtained by estimating each moment of the distribution and solving for the parameters.

Now it's time for us to consider the distribution of these estimators themselves.

Def: Given a random sample $\mathbf{X}$, let T be a function of $\mathbf{X}$ and possibly θ. Since $\mathbf{X}$ is random, them $T(\mathbf{X},\theta)$ is also a random variable. It's distribution is called the **sampling distribution of T**

Note: in the special case when T does not depend on θ, then T will be a **statistic**. If the statistic is used to estimate a parameter θ, we can use the sampling distribution of the statistic to assess the probability that the estimator is close to θ.

Ex 1: Let $X_1,\dots,X_n \sim N(\mu,\sigma^2)$. We define two functions, R and T:

$$R(\mathbf{X}) = \frac{1}{n}\sum X_i = \bar{X} \qquad T(\mathbf{X},\mu,\sigma^2) = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

Note that R is a statistic, while T is not.

Since the X_i are Normally distributed, then $\bar{X}$ is also Normal. In particular, the sampling distribution of $\bar{X}$ is $N(\mu,\sigma^2/n)$.

By location-scale transformations, the variable $T(\mathbf{X},\mu,\sigma^2)$ is Normally distributed, with mean of 0 and variance 1. Importantly, it turns out that the distribution of T does not depend on the values of the potentially unknown parameters μ and σ^2.

Ex 2: Suppose $X_1,\dots,X_n \sim \text{Unif}(0,\theta)$. What are some possible estimators we could use to estimate θ? Find at least 3.

Ex 3: Suppose $X_1,\dots,X_n \sim \text{Unif}(0,\theta)$. The MLE $\hat{\theta}_{MLE}$ for θ is $\max\{X_i\}$, which has density function:

$$f(x) = \frac{n}{\theta^n}x^{n-1}$$

and so $\frac{\hat{\theta}_{MLE}}{\theta} \sim \text{Beta}(n,1)$. The mean and variance of $\hat{\theta}_{MLE}$ are

$$E[\hat{\theta}_{MLE}] = \frac{n}{n+1}\theta \qquad \text{var}(\hat{\theta}_{MLE}) = \frac{n\theta^2}{(n+1)^2(n+2)}$$

On the other hand, the MoM estimator for θ is $\hat{\theta}_{MoM} = 2\bar{X}$, whose distribution doesn't have a particularly nice form (it is the density for a sum of iid uniform variables). However, by the Central Limit Theorem, if n is relatively large (in this case, $n \geq 5$ is probably fine), then $\hat{\theta}_{MoM}$ is approximately Normal. And in any case, with mean and variance

$$E[\hat{\theta}_{MoM}] = \theta \qquad \text{var}(\hat{\theta}_{MoM}) = \frac{\theta^2}{3n}$$

Ex 4: For each of the MLE and MoM estimators, what sample size is necessary to ensure the standard deviation of the estimator is less than 1% of the value of θ?

Solution. For $\hat{\theta}_{MoM}$, $n \geq \frac{1}{0.01^2\sqrt{3}} \approx 5774$. For $\hat{\theta}_{MLE}$, $n \geq 99$. ■

Ex 5: Use R to simulate samples from $\text{Unif}(0,10)$. Compare the sampling distributions of the two (or more) estimators.