STA 623 – Fall 2010 – Dr. Charnigo

Section 5.2: Sums of Random Variables from a Random Sample

Statistics. Suppose that $X_1, \ldots, X_n$ are iid. Let $T(x_1, \ldots, x_n)$ be a(n appropriately measurable) function from $\mathbb{R}^n$ to $\mathbb{R}^k$ for some positive integer k. We refer to the random variable or vector $T(X_1, \ldots, X_n)$ as a statistic and to its probabilistic behavior as a sampling distribution.

Sample mean. The sample mean $\bar{X} := n^{-1} \sum_{i=1}^n X_i$ is one of the most familiar statistics. Presuming that $X_1, \ldots, X_n$ have common mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in (0, \infty)$, we readily calculate that

$$E[\bar{X}] = n^{-1} \sum_{i=1}^n E[X_i] = n^{-1} \sum_{i=1}^n \mu = \mu$$

and

$$Var[\bar{X}] = n^{-2} \sum_{i=1}^n Var[X_i] = n^{-2} \sum_{i=1}^n \sigma^2 = \sigma^2/n.$$

Since each of $\bar{X}$ and μ is called a mean, referring to the former as a sample mean and to the latter as a population mean can reduce confusion. The first result above also shows that the sample mean is an unbiased estimator of the population mean, in the sense that the expected value of the estimator equals the quantity being estimated. In particular, the sample mean has no systematic tendency, over repeated sampling, either to overestimate or to underestimate the population mean.

Neither result above requires $X_1, \ldots, X_n$ to be normally distributed. On the other hand, if $X_1, \ldots, X_n$ are normally distributed, then so is $\bar{X}$. Astonishingly, $\bar{X}$ is approximately normally distributed even if $X_1, \ldots, X_n$ are not, the approximation becoming better and better as n increases. Formally, for any real number x we have

$$\lim_{n\to\infty} P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq x\right) = \int_{-\infty}^x (2\pi)^{-1/2} \exp[-t^2/2]\, dt.$$

This kind of limiting behavior — that is, convergence of cumulative distribution functions — is referred to as convergence in distribution or, synonymously, convergence in law. The particular result shown above is called the Central Limit Theorem. You will see a proof of the Central Limit Theorem in STA 606. The proof uses moment generating functions.

Sample variance. The sample variance $S^2 := (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$ and sample standard deviation $S := \sqrt{S^2}$ are also well known. Put $c_1 := (n-1)/n$, $c_2 := -1/n$, $c_3 := -1/n, \ldots, c_n := -1/n$. We have

$$\begin{aligned}
E[S^2] &= (n-1)^{-1}\sum_{i=1}^n E[(X_i - \bar{X})^2] \\
&= (n-1)^{-1} n E[(X_1 - \bar{X})^2] \\
&= (n-1)^{-1} n E[(\sum_{j=1}^n c_j X_j)^2] \\
&= (n-1)^{-1} n E[\sum_{j=1}^n \sum_{k=1}^n c_j c_k X_j X_k] \\
&= (n-1)^{-1} n \sum_{j=1}^n \sum_{k=1}^n c_j c_k E[X_j X_k] \\
&= (n-1)^{-1} n \sum_{j=1}^n \sum_{k=1}^n c_j c_k (\mu^2 + \sigma^2 1_{\{j=k\}}) \\
&= (n-1)^{-1} n \{\sum_{j=1}^n \sum_{k=1}^n c_j c_k \mu^2 + \sum_{j=1}^n c_j^2 \sigma^2\} \\
&= (n-1)^{-1} n \{\mu^2 \sum_{j=1}^n c_j \sum_{k=1}^n c_k + \sigma^2[(n-1)^2 + (n-1)]/n^2\} \\
&= (n-1)^{-1} n \sigma^2 (n-1) n / n^2 \\
&= \sigma^2.
\end{aligned}$$

Your textbook authors give a much shorter proof of this result by appealing to the fact that $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$. Understanding the longer proof is worthwhile, however, since it illustrates a general technique for calculating the expected value of any positive power of any linear combination of $X_1, \ldots, X_n$.

Presuming that $X_1, \ldots, X_n$ have finite fourth moment, one can also show that

$$Var[S^2] = n^{-1}\{E[(X_1 - \mu)^4] - (n-3)Var[X_1]^2/(n-1)\}.$$

If $X_1, \ldots, X_n$ are normally distributed, then $E[(X_1 - \mu)^4] = 3\sigma^4$ and

$$Var[S^2] = 2\sigma^4/(n-1).$$

Techniques for finding the distribution of a sum. You have already seen that moment generating functions may be useful for finding the distribution of $X_1 + \cdots + X_n$. However, sometimes this approach fails because either: (i) the moment generating function of $X_1 + \cdots + X_n$ is not recognizable; or, (ii) $X_1, \ldots, X_n$ do not have a finite moment generating function in a neighborhood of 0.

An alternative approach that sometimes works is to apply the convolution formula

$$f_{X_1+X_2}(x) = \sum_{w \in \mathbb{Z}} f_{X_1}(w) f_{X_2}(x-w)$$

for integer-valued discrete random variables or

$$f_{X_1+X_2}(x) = \int_{\mathbb{R}} f_{X_1}(w) f_{X_2}(x-w)\; dw$$

for real-valued continuous random variables, perhaps followed by mathematical induction. The convolution formula does not require that X_1 and X_2 be identically distributed but does require that they be independent. Your textbook authors prove the continuous case via the bivariate transformation formula. A proof of the discrete case emerges from the law of total probability with $A := \{X_1 + X_2 = x\}$ and $B_w := \{X_1 = w\}$ for $w \in \mathbb{Z}$,

$$P(A) = \sum_{w \in \mathbb{Z}} P(A|B_w)P(B_w)$$

or

$$\begin{aligned} P(X_1 + X_2 = x) &= \sum_{w \in \mathbb{Z}} P(X_1 + X_2 = x | X_1 = w) P(X_1 = w) \\ &= \sum_{w \in \mathbb{Z}} P(X_2 = x - w | X_1 = w) P(X_1 = w) \\ &= \sum_{w \in \mathbb{Z}} P(X_2 = x - w) P(X_1 = w). \end{aligned}$$