There are many other estimators of the population cdf $F(x)$. If it is known that $F(\cdot)$ is continuous, one might want to use an estimator that is also continuous, and in practice this can be accomplished by "smoothing" the sample cdf.

As a generalization of the sample cdf defined above for univariate random variables, consider the following model. Assume

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \overset{\text{ident}}{\sim} F(\cdot, \cdot).$$

Definition 7 The *bivariate sample cdf*, denoted by $F_n(\cdot,\cdot)$, is defined by $F_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)\, I_{(-\infty,y]}(Y_i)$. ////

Note that $F_n(x,y)$ is the proportion of the (X_i, Y_i) pairs lying southwest of the point (x,y) in the plane.

Again the mean, variance, and distribution of $F_n(x,y)$ can be readily obtained. For example, $F_n(x,y) \sim$ "scaled" bin$(n, p = F(x,y))$. See the Problems for some details.

3.2 Sample Mean

Recall that the sample mean, which is also the mean of the sample distribution, was defined to be $\overline{X}_n = \overline{X} = \frac{1}{n}\sum_1^n X_i$ for the sample $X_1, \ldots, X_n$ of random variables.

PROPERTIES OF THE SAMPLE MEAN

(i) Assume $X_1, \ldots, X_n \sim (\mu)$ (says all the X_i's are distributed with common mean μ). Then $\mathcal{E}[\overline{X}_n] = \mu$; that is, under minimal restrictions, the expected value of the sample mean equals the population mean.

(ii) $\text{var}[\overline{X}_n] = \text{cov}\left[\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right] = \frac{1}{n^2}\sum_i\sum_j \text{cov}[X_i, X_j]$

where the only assumption so far is the existance of the covariances. We would like $\text{var}[\overline{X}_n] \longrightarrow 0$ as $n \longrightarrow \infty$. One can easily see what it takes; the n^2 of the denominator must go to infinity faster than the sum of the n^2 terms in the double sum of the numerator.

N.B. $X_1, \ldots, X_n \sim (\mu, \sigma^2)$ and X_i, X_j uncorrelated for $i \neq j$ implies $\text{var}[\overline{X}_n] = \frac{\sigma^2}{n}$. See the Problems for various examples.

(iii) Under the iid model (as well as other models) theoretically the exact distribution of $\overline{X}_n$ is known or can be obtained. For example, recall the distribution of the sum of iid Bernoullis, binomials, Poissons, geometrics, negative binomials, normals, uniforms, exponentials, gammas, and Cauchys. For some other cases the distribution of the sum is not easy to find and the answer is not necessarily simple; e.g., hypergeometrics, Weibulls, Gumbels, etc. It is such cases that motivate asymptotics to be studied in Section 5.

EXAMPLE 3 Assume $X_1, \ldots, X_n \overset{iid}{\sim} \exp(\beta)$. Then $\sum_1^n X_i \sim$ gamma (n, β); so, for $z < 0$, $F_{\overline{X}_n}(z) = P\left[\overline{X}_n \le z\right] = P\left[\sum_1^n X_i \le nz\right] = \int_0^{nz} \frac{1}{\Gamma(n)} \left(\frac{1}{\beta}\right)^n x^{n-1} e^{-x/\beta} dx$ and hence $f_{\overline{X}_n}(z)$ $= \frac{n}{\Gamma(n)} \left(\frac{1}{\beta}\right)^n (nz)^{n-1} e^{-\frac{nz}{\beta}} I_{(0,\infty)}(z)$; that is, $\overline{X}_n \sim$ gamma $(n, \beta/n)$. ////

3.3 Sample Variance

Recall that the sample variance was defined to be $S_n^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ for random variables $X_1, \ldots, X_n$.

PROPERTIES OF THE SAMPLE VARIANCE

(i) S_n^2 is a function of the $n-1$ "residuals":

$$X_1 - \overline{X}_n, X_2 - \overline{X}_n, \ldots, X_{n-1} - \overline{X}_n$$

noting that $X_n - \overline{X}_n$ is a function of the first $n-1$ residuals since

$$\sum_{i=1}^n (X_i - \overline{X}_n) = 0.$$

This property says that S_n^2 is a function of $n-1$ bits of information which partially justifies the divisor of $n-1$.

(ii) $S_n^2 = \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2.$

PROOF

$$\begin{aligned}
\tfrac{1}{2n(n-1)} \sum\sum (X_i - X_j)^2 &= \tfrac{1}{2n(n-1)} \sum\sum ((X_i - \overline{X}) - (X_j - \overline{X}))^2 \\
&= \tfrac{1}{2n(n-1)} \left\{\sum\sum \left((X_i - \overline{X})^2 - 2(X_i - \overline{X})(X_j - \overline{X}) + (X_j - \overline{X})^2\right)\right\} \\
&= \tfrac{1}{2n(n-1)} \left\{n \sum (X_i - \overline{X})^2 - \text{ zero } + n \sum (X_j - \overline{X})^2\right\} \\
&= \tfrac{\sum (X_i - \overline{X})^2}{n-1}.
\end{aligned}$$

////

Properties (i) and (ii) are strictly algebraic and require no assumptions about the background random variables $X_1, \ldots, X_n$. Note that (ii) says that S_n^2 is a function of, say, the $n-1$ differences $X_2 - X_1, \ldots, X_n - X_1$, noting that $X_i - X_j = (X_i - X_1) - (X_j - X_1)$. The sample variance supposedly measures the "spread" of the X_i's, so it's reasonable that such "spread" can be characterized by the difference between pairs of the X_i's.

THEOREM 3.3 *Under the iid model* $\mathcal{E}[S_n^2] = \sigma^2$ *and* $\mathrm{var}[S_n^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)$.

PROOF

$$\begin{aligned}
\mathcal{E}[S_n^2] &= \tfrac{1}{2n(n-1)} \sum\sum_{i \neq j} \mathcal{E}\left[((X_i - \mu) - (X_j - \mu))^2\right] \\
&= \tfrac{1}{2n(n-1)} \sum\sum_{i \neq j} \mathcal{E}\left[(X_i - \mu)^2 - 2(X_i - \mu)(X_j - \mu) + (X_j - \mu)^2\right] \\
&= \tfrac{1}{2n(n-1)} \sum\sum_{i \neq j} \left[\sigma^2 - 0 + \sigma^2\right] \\
&= \sigma^2 \text{ (Note that independence was not used, only zero corrrelation and common } \sigma^2\text{.)} \\
\mathrm{var}[S_n^2] &= \tfrac{1}{[2n(n-1)]^2} \mathrm{var}\left[\sum_i \sum_j (X_i - X_j)^2\right] = \tfrac{1}{[2n(n-1)]^2} \mathrm{cov}\left[\sum_i \sum_j (X_i - X_j)^2, \sum_\alpha \sum_\beta (X_\alpha - X_\beta)^2\right] \\
&= \tfrac{1}{[2n(n-1)]^2} \sum_i \sum_j \sum_\alpha \sum_\beta \mathrm{cov}\left[(X_i - X_j)^2, (X_\alpha - X_\beta)^2\right]
\end{aligned}$$

Now partition the n^4 terms into the cases:

indices		# of such terms	value of term
$i=j=\alpha=\beta$ (all alike)		n	zero
$i=j=\alpha\neq\beta$ (× 4) (three alike)		$4n(n-1)$	zero
$i=j\neq\alpha=\beta$	two pairs alike	$n(n-1)$	zero
$i=\alpha\neq j=\beta$		$n(n-1)$	$\text{cov}[(X_1-X_2)^2,(X_1-X_2)^2]$
$i=\beta\neq j=\alpha$		$n(n-1)$	$\text{cov}[(X_1-X_2)^2,(X_1-X_2)^2]$
$i=j\neq\ \beta,\ \alpha\neq\beta$	one pair alike, the rest different	$n(n-1)(n-2)$	zero
$i=\alpha$ etc.		$n(n-1)(n-2)$	$\text{cov}[(X_1-X_2)^2,(X_1-X_3)^2]$
$i=\beta$ etc.		$n(n-1)(n-2)$	$\text{cov}[(X_1-X_2)^2,(X_1-X_3)^2]$
$j=\alpha$ etc.		$n(n-1)(n-2)$	$\text{cov}[(X_1-X_2)^2,(X_1-X_3)^2]$
$j=\beta$ etc.		$n(n-1)(n-2)$	$\text{cov}[(X_1-X_2)^2,(X_1-X_3)^2]$
$\alpha=\beta$ etc.		$n(n-1)(n-2)$	zero
all different		$n(n-1)(n-2)(n-3)$	zero (by indep.)

$$
\begin{aligned}
&= \tfrac{1}{[2n(n-1)]^2}\left\{2n(n-1)\text{var}\left[(X_2-X_1)^2\right]+4n(n-1)(n-2)\ \text{cov}\left[(X_1-X_2)^2,(X_1-X_3)^2\right]\right\}\\
&= \tfrac{1}{[2n(n-1)]^2}\left\{2n(n-1)(2\mu_4+2\sigma^4)+4n(n-1)(n-2)(\mu_4-\sigma^4)\right\}\\
&= \tfrac{1}{n}\left(\mu_4-\tfrac{n-3}{n-1}\sigma^4\right),\ \text{noting}\\
&\quad \text{var}\left[(X_2-X_1)^2\right]=\\
&\quad \text{var}\left[(X_2-\mu)^2-2(X_1-\mu)(X_2-\mu)+(X_1-\mu)^2\right]=\\
&\quad \text{var}\left[(X_2-\mu)^2\right]-4\underbrace{\text{cov}\left[(X_2-\mu)^2,(X_1-\mu)(X_2-\mu)\right]}_{=0}+2\underbrace{\text{cov}\left[(X_2-\mu)^2,(X_1-\mu)^2\right]}_{=0}\\
&\quad +4\ \text{var}\left[(X_1-\mu)(X_2-\mu)\right]-4\underbrace{\text{cov}\left[(X_1-\mu)(X_2-\mu),(X_1-\mu)^2\right]}_{=0}+\text{var}\left[(X_1-\mu)^2\right]=\\
&\quad 2\ \text{var}\left[(X_1-\mu)^2\right]+4\ \text{var}\left[(X_1-\mu)(X_2-\mu)\right]=\\
&\quad 2\ \mathcal{E}\left[(X_1-\mu)^4\right]-2\ \mathcal{E}^2\left[(X_1-\mu)^2\right]+4\ \mathcal{E}\left[(X_1-\mu)^2(X_2-\mu)^2\right]-4\ \mathcal{E}^2\left[(X_1-\mu)(X_2-\mu)\right]=\\
&\quad 2\ \mu_4 \qquad -2\ \sigma^4 \qquad +4\ \sigma^4 \qquad -\ \text{zero}, \qquad \text{and}\\
&\quad \text{cov}\left[(X_1-X_2)^2,(X_1-X_3)^2\right]=\text{cov}\left[(X_1-\mu)^2-2(X_1-\mu)(X_2-\mu)\right.\\
&\quad \left.+(X_2-\mu)^2;(X_1-\mu)^2-2(X_1-\mu)(X_3-\mu)+(X_3-\mu)^2\right]=\\
&\quad \text{var}\left[(X_1-\mu)^2\right] \qquad +\ \text{eight zeros} \qquad = \qquad \mu_4-\sigma^4 \qquad ////
\end{aligned}
$$

This theorem says, under the assumed model, that the expected value of the sample variance is the population variance and the variance of the sample variance is "small" when n is "large," indicating that the sample variance is not a bad estimator of the population variance.

If a distribution is assumed for $X_1,\ldots,X_n$, then, theoretically at least, one should be able to derive the distribution of S_n^2. With the exception of the normal distribution covered in Section 4, such derivation is not necessarily nice so we forego exact distributional results on S_n^2 and await the asymptotics of Section 5.

3.4 Covariance of Sample Mean and Sample Variance

Since $\overline{X}_n$ and S_n^2 are both functions of the same $X_1, \ldots, X_n$, one expects them to be dependent, so it might be somewhat surprising that the covariance between the two is as simple as it is.

THEOREM 3.4 *Under the iid model,* $\text{cov}\left[\overline{X}_n, S_n^2\right] = \frac{\mu_3}{n}$.

PROOF

$$
\begin{aligned}
&\text{cov}\left[\overline{X}_n, S_n^2\right] \\
&= \text{cov}\left[\overline{X}_n - \mu, \tfrac{1}{2n(n-1)} \sum \sum (X_i - X_j)^2\right] \\
&= \tfrac{1}{2n^2(n-1)} \text{cov}\left[\textstyle\sum_i (X_i - \mu),\ \sum_\alpha \sum_\beta \left((X_\alpha - \mu) - (X_\beta - \mu)\right)^2\right] \\
&= \tfrac{1}{2n^2(n-1)} \textstyle\sum_i \sum_\alpha \sum_\beta \text{cov}\left[(X_i - \mu), \left((X_\alpha - \mu) - (X_\beta - \mu)\right)^2\right]
\end{aligned}
$$

Now partition the n^3 terms as was done in the proof of the previous theorem and μ_3/n results. ////

Note that $\overline{X}_n$ and S_n^2 are uncorrelated if and only if $\mu_3 = 0$; further, $\mu_3 = 0$ for any symmetric distribution for which the third moment exists. We shall see in Section 4 that $\overline{X}_n$ and S_n^2 are actually independent for the normal distribution.

3.5 Higher Order Sample Moments

The rth raw sample moment of the sample $X_1, \ldots, X_n$ was defined to be

$$M_r' = \frac{1}{n} \sum_{i=1}^{n} X_i^r.$$

Since M_r' is the sample mean of $X_1^r, \ldots, X_n^r$ it has many of the same properties as the sample mean. For instance, $\mathcal{E}[M_r'] = \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}[X_i^r] = \mu_r'$, the rth raw population mean, under any model with $\mathcal{E}[X_i^r] = \mu_r'$ for $i = 1, \ldots, n$. Also, under an iid model,

$$\text{var}\left[M_r'\right] = \text{var}\left[X_i^r\right]/n = \left[\mu_{2r}' - (\mu_r')^2\right]/n.$$

Exact results for higher order *central* sample moments are not so nice. The same is true for the sample skewness, sample kurtosis, and sample coefficient of variation. All of these involve ratios of statistics, and we saw in Chapter V that even the first and second moments of ratios can be messy. We will, however, return to these statistics when we study asymptotics in Section 5.