Stat 425 HW6

Fritz Scholz

1. Write a function

$$\texttt{Estimate.compare} = \texttt{function(m} = \texttt{10,n} = \texttt{10,Nsim} = \texttt{1000,Delta} = \texttt{2,}$$
$$\texttt{dist} = \texttt{rnorm,parm} = \texttt{c(0,1))}\{...\}$$

that generates independent samples $X_, \ldots, X_m \sim F(x)$ and $Y_1, \ldots, Y_n \sim F(x-\Delta)$ and computes the estimates $\bar{\Delta} = \bar{Y} - \bar{X}$ and $\hat{\Delta} = \text{med}_{i,j}(Y_j - X_i)$ and repeats this process `Nsim` times and collects the estimates in vectors `Delta.bar` and `Delta.hat`, respectively. Here the input argument `dist` indicates which F is to be used. This builds on what you learned from the previous assignment. Inside the function body of `Estimate.compare` you may want to use the function `Using.dist` given in the solution to HW5. Of course, you have to make sure that `Using.dist` exists inside your R work space.

The output of `Estimate.compare` should be the mean, median and variance of the two generated vectors of length `Nsim`, i.e., the estimated mean, median and variance of the respective estimator distributions. We know that the variance of the $\bar{\Delta}$ sampling distribution is

$$\text{var}(\bar{\Delta}) = \text{var}(\bar{Y} - \bar{X}) = \text{var}(\bar{Y}) + \text{var}(\bar{X}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \sigma^2 \, \frac{m+n}{mn}$$

where σ^2 is the variance of $F(x)$ and thus also of $F(x-\Delta)$. We see that $1/\text{var}(\bar{\Delta})$ is proportional to $mn/(m+n)$. Illustrate this fact by writing a function

$$\texttt{recip.var.plot} = \texttt{function(Nsim} = \texttt{1000,mvec} = \texttt{c(5,10,20,30,50,100),}$$
$$\texttt{nvec} = \texttt{c(5,10,20,30,50,100),dist} = \texttt{rnorm,}$$
$$\texttt{parm} = \texttt{c(0,1),dist.name} = \texttt{"normal")}\{\ldots\}$$

that calls `Estimate.compare` for several different m, n, i.e., $m = \texttt{mvec[i]}$ and $n = \texttt{nvec[i]}, \texttt{i} = 1,2,\ldots$, and which plots the reciprocal of the estimated variances, i.e., the estimated values of $1/\text{var}(\bar{\Delta})$, against the respective values of $mn/(m+n)$. The point pattern should look roughly linear. Fit a line through this point pattern such that it goes through the origin (i.e., with intercept zero) using `out.ls1=lsfit(...,intercept=F)`. The component `out.ls1$coef` will give you the slope of that line. Does the slope of this linear pattern make sense? Once you have accomplished this, add the corresponding points for the estimates of $1/\text{var}(\hat{\Delta})$ in relation to $mn/(m+n)$ and fit a similar line to them. Does this suggest that $\text{var}(\hat{\Delta})$ is proportional to $(m+n)/mn$? Make sure that all 12 points show in the plot. Annotate your plot such that the sampled distribution is shown (that's is where you use the argument value for `dist.name`). In the annotation also show the ratio of slopes (slope of the $1/\text{var}(\hat{\Delta})$ pattern over the slope of the $1/\text{var}(\bar{\Delta})$ pattern) and also show those slopes separately. Further, add a legend explaining which points/lines belong to which type of estimate. Using an appropriate choice for `M` you may want to experiment with:

```
text(0,.9*M,substitute("slope for "~1/var(hat(Delta))==xx,
     list(xx=format(signif(out.ls2$coef,3))))),adj=0)
text(0,.85*M,substitute("slope for "~1/var(bar(Delta))==xx,
     list(xx=format(signif(out.ls1$coef,3))))),adj=0)
legend(0,.8*M,c(expression(1/var(bar(Delta))),
     expression(1/var(hat(Delta)))),lty=1:2,
     col=c("black","blue"),pch=c(1,16),bty="n")
```

We have shown above why to expect proportionality with respect to $mn/(m+n)$ in the case of $1/\text{var}(\bar{\Delta})$ and based on the simulation evidence we will accept the same type of proportionality for $1/\text{var}(\hat{\Delta})$.
If you were to match $\text{var}(\hat{\Delta}_{m.n}) = \text{var}(\bar{\Delta}_{m',n'})$ by proper choice of sample sizes $m = n$ and $m' = n' = \rho \times m$, what concept would ρ resonate with and how does it relate to the ratio of slopes that annotates your plot. We have seen two instances of this concept already.
By changing `dist`, `parm`, `dist.name` make such plots for `dist = rnorm`, `dist = runif`, `dist = rlogis`, and `dist = rexp`.

What is expected:

1. The code for `Estimate.compare`.
2. The code for `recip.var.plot`.
3. The 4 plots produced by `recip.var.plot` for the 4 indicated distributions.
4. Discussion of the resonance of ρ and its relation to the slope ratio.
5. Does the slope for $1/\text{var}(\bar{\Delta})$ in the normal distribution plot make sense? Explain. Try to extend that insight to the other distributions. Take as given: The variance of the logistic distribution with location 0 and scale 1 is $\pi^2/3$, the variance of the $U(0,1)$ distribution is $1/12$, and the variance of the exponential distribution with mean 1 is 1.