Celebrating 40 Years of the GSI, 1981-2021

Variance Estimation for a Univariate Normal Distribution

Mark Brown

Department of Statistics,
Columbia University,
New York, NY, USA

Dedicated to the memory of Theofilos Cacoullos

An old familiar problem

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, μ, σ^2 unknown. Define

$$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

and

$$\hat{\sigma}_n^2(c) = \frac{S^2}{n - 1 + c},$$

for $-(n-1) < c < \infty$.

- The minimum variance unbiased estimator $\sigma_n^2(0)$ appears to be the most popular and widely used estimator.
- This is true despite the fact (well-known) that the maximum likelihood estimator, $\hat{\sigma}_n^2(1) = S^2/n$, has smaller mean square error for all σ^2.

An old familiar problem (continued)

- The estimator in the class $\hat{\sigma}_n^2(c)$ with smallest mean square error is $\hat{\sigma}_n^2(2) = S^2/(n+1)$. The mle is sometimes used, but not apparently $\hat{\sigma}_n^2(2)$. These are very well known results and are found in numerous textbooks in statistics.

- The preference for $\hat{\sigma}_n^2(0)$ would seem to suggest that unbiasedness is a highly valued attribute. Nevertheless, $\hat{\sigma}_n(0)$, known as the sample standard deviation, is the estimator of choice for σ, but is biased.

- It is a routine exercise to find a constant multiple of $\hat{\sigma}_n(0)$ which is unbiased, but there do not appear to be advocates for this approach.

An old familiar problem (continued)

This seems to me to be a basic logical problem. Why is unbiasedness of importance for estimating σ^2, but not for σ? When I ask people, I get answer like:

1. *Who cares?* For n moderately large the difference between the estimators are negligible.
2. *Data is not really exactly normal anyway.* All of these estimators are similar and their differences are minuscule compared to the problem of departures from normality. We should instead use robust estimators of scale.
3. *One should use a Bayesian approach* to choose the appropriate multiple of σ^2.

Choice of Loss Function

- The use of mean square error is widely acknowledged, (going back to Gauss), to be primarily motivated by mathematical convenience. In this example, we should ask, "what is the purpose of estimating σ^2?".

- One answer is that from the data we will estimate various probabilities of events under $\mathcal{N}(\mu, \sigma^2)$ by the corresponding probabilities under $\mathcal{N}(\bar{X}, \hat{\sigma}^2)$, where $\hat{\sigma}^2$ is the estimated variance, and $\bar{X}$ the sample mean.

- One measure of this loss would be:

$$\ell(\hat{\sigma}^2, \sigma^2) = \sup_B \left| P\left(\mathcal{N}(\mu, \hat{\sigma}^2) \in B\right) - P\left(\mathcal{N}(\mu, \sigma^2) \in B\right) \right| .$$

Choice of Loss Function (continued)

This loss is independent of μ and satisfies:

$$\ell(\hat{\sigma}^2, \sigma^2) = g\left(\frac{\hat{\sigma}^2}{\sigma^2}\right), \tag{1}$$

where

$$g(r) = g\left(\frac{1}{r}\right) \text{ for } 0 < r < \infty; \tag{2}$$

$$g \text{ is strictly increasing on } [1, \infty). \tag{3}$$

The g function corresponding to total variation distance is given by,

$$g_v(r) = 2\left|\Phi\left(\left(\frac{r\log r}{r-1}\right)^{1/2}\right) - \Phi\left(\left(\frac{\log r}{r-1}\right)^{1/2}\right)\right|,$$

where Φ is the standard normal cdf.

Log Symmetric Loss

- A loss function with properties (1)–(3) will be called a log symmetric loss function. This class coincides with the class of loss functions which are increasing functions of total variation distance,

$$d(\sigma_1^2, \sigma_2^2) = y\left(g_v\left(\frac{\sigma_1^2}{\sigma_2^2}\right)\right),$$

  where y is strictly increasing on $[0, 1)$.

- Other examples of such loss functions are $g_1(r) = |\log(r)|$ and $g_2(r) = \left[1 - \left(\frac{2r^{1/2}}{1+r}\right)^{1/2}\right]^{1/2}$. g_1 is known as arc length distance and g_2 is Hellinger distance.

Log Symmetric Loss (continued)

- Consider the class $\{\hat{\sigma}^2(c), -(n-1) < c < \infty\}$ defined above. Denote by $L(c)$ a random variable whose distribution is that of the loss incurred by the use of $\hat{\sigma}^2(c)$ under a fixed but arbitrary log symmetric loss function. Then,

$$P(L(c) \leq x)$$

  is non-increasing in $c \geq 0$ for all x, and strictly decreasing in $c \geq 0$ for any $x \in (g(1^+), g(\infty))$.

- The stochastic dominance property is considerably stronger than $\mathbb{E}L(c)$ increasing in $c \geq 0$, when this expectation exists.

Log Symmetric Loss (continued)

- We conclude that under log symmetric loss, for $0 \leq c_1 < c_2$, that $\hat{\sigma}^2(c_1)$ is a better estimator than $\hat{\sigma}^2(c_2)$. Thus the minimum variance unbiased estimator, $\hat{\sigma}^2(0)$, is better than the mle, $\sigma^2(1)$, which in turn is better than the minimum mean square error estimator in this class, $\hat{\sigma}^2(2)$. The conclusion provides support for conventional wisdom.

- The above stochastic monotonicity does not extend to $c \leq 0$. Under a log symmetric function, no $\hat{\sigma}^2(c)$ with $-(n-1) < c \leq 0$, will stochastically dominate another member of this subclass. Within this subclass we need to discriminate among estimators by the weaker criteria of expected loss.

Log Symmetric Loss (continued)

- An asymptotic analysis suggests that $c = -2/3$ gives minimal asymptotic risk for a wide class of log symmetric loss functions. Thus,

$$\hat{\sigma}^2\left(-\frac{2}{3}\right) = \frac{S^2}{n - \frac{5}{3}},$$

  would be the estimate of choice under this criteria.

- Comparing $\hat{\sigma}^2(0)$ to $\hat{\sigma}^2(-2/3)$, under the total variation distance,

$$\begin{aligned} \mathbb{E}L(0) - \mathbb{E}L\left(-\frac{2}{3}\right) &= \frac{1}{9\pi}\sqrt{\frac{2}{e}}(n-1)^{-3/2} + o\left(n^{-2}\right) \\ &\sim .03034(n-1)^{-3/2}. \end{aligned}$$

Linear Models

- Consider $X^{n\times 1} \sim \mathcal{N}(\mu, \sigma^2\Sigma)$, with Σ a known positive definite symmetric matrix, σ^2 unknown, and μ known to lie in $\mathcal{M}$, a p-dimensional subspace of $\mathbb{R}^n$.
- Here the analog of $\hat{\sigma}^2(c)$ is $S^2/(n-p+c)$, where $S^2 = ||X - P_{\mathcal{M}}X||^2$, and $P_{\mathcal{M}}X$ is the projection of X on $\mathcal{M}$ under the inner product $(x, y) = x^\top \Sigma^{-1} y$.
- Once again $L(c)$ is stochastically increasing in $c \geq 0$, and the minimum variance unbiased estimator, $\sigma^2(0)$, stochastically dominates the mle, $\sigma^2(p)$.

Thank you!