

Concentrated Likelihood Functions, Normal Equations, and Properties of Maximum Likelihood: Lecture XXII

Charles B. Moss

October 12, 2026

- 1 Distribution of the Sum of Normals - the Sample Mean
- 2 Concentrated Likelihood Functions
- 3 Normal Equations
- 4 Properties of Maximum Likelihood Estimators
 - Consistency of Maximum Likelihood
 - Asymptotic Normality

Distribution of the Sum of Normals

- Suppose we start with a random variable Y defined as

$$Y = \sum_{i=1}^N c_i x_i \quad (1)$$

- Next, suppose that the x_i s were independent and identically distributed normal random variables. Assume that the moment generating function for these variables is

$$M_{x_i}(t) = \exp\left(\mu t - \frac{\sigma^2 t^2}{2}\right) \quad (2)$$

- This would imply that the moment generating function for Y is

$$M_Y(t) = \prod_{i=1}^N \exp\left(c_i \mu t - \frac{c_i \sigma^2 t^2}{2}\right)$$
$$\Rightarrow Y \sim N\left[\mu \sum_{i=1}^N c_i, \sigma^2 \sum_{i=1}^N c_i^2\right] \quad (3)$$

- In the case of the sample mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \sim N\left[\mu, \frac{\sigma^2}{N}\right] \quad (4)$$

From Last Lecture

- From last lecture we rewrote

$$\begin{aligned} f(x|\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2}\right) \times \\ &\quad \exp\left[-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}\right] \end{aligned} \quad (5)$$

- We decompose the result into a distribution of $\bar{x}$ and the remaining (mean - corrected) distribution in the sample.
- The distribution of $\bar{x}$ is

$$q(T(x) | \theta) = \frac{1}{\left(2\pi \frac{\sigma^2}{n}\right)^{\frac{1}{2}}} \exp \left[-\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right]. \quad (6)$$

- Dividing Equation 5 by Equation 6 yields

$$g(\{x_i\}_{i=1}^n | \sigma^2) = \frac{1}{n^{\frac{1}{2}} (2\pi\sigma^2)^{\frac{n-1}{2}}} \exp \left[-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2} \right]. \quad (7)$$

Concentrated Likelihood Functions

- In the last lecture I introduced the concept of maximum likelihood using a known variance normal distribution of unknown mean. Undoubtedly this framework is fictional. Even if we know that the distribution is normal, it would be a rare event to know the distribution of the variance.
- What I would like to do now is to demonstrate how maximum likelihood functions may be applied by demonstrating how the variance can be concentrated out of the standard normal estimation.
 - The more general form of the normal likelihood function can be written as

$$L(X|\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(X_i - \mu)^2}{2\sigma^2} \right] \quad (8)$$

- Continued

- Ignoring the constants, the natural logarithm of the likelihood function can be written as

$$\ln(L) = -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \quad (9)$$

- This expression can be solved for the optimal choice of σ^2 by differentiating with respect to σ^2

$$\begin{aligned} \frac{\partial \ln(L)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (X_i - \mu)^2 = 0 \\ \Rightarrow -n\sigma^2 + \sum_{i=1}^n (X_i - \mu)^2 &= 0 \\ \Rightarrow \hat{\sigma}_{MLE}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \end{aligned} \quad (10)$$

- Continued

- Substituting this result into the original logarithmic likelihood yields

$$\begin{aligned}\ln(L) &= -\frac{n}{2} \ln \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right) - \frac{1}{2 \frac{1}{n} \sum_{j=1}^n (X_j - \mu)^2} \sum_{i=1}^n (X_i - \mu)^2 \\ &= -\frac{n}{2} \ln \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right) - \frac{n}{2}\end{aligned}\tag{11}$$

- Intuitively, the maximum likelihood estimate of μ is that value that minimizes the mean square error of the estimator. Thus, the least squares estimate of the mean of a normal distribution is the same as the maximum likelihood estimator under the assumption that the sample is independently and identically distributed.

Maximizing the Concentrated Likelihood Function

- Starting with the result of Equation 11

$$\begin{aligned}\ln(L) &= -\frac{n}{2} \ln \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right) - \frac{n}{2} \\ &= -\frac{n}{2} \ln \left(\frac{1}{n} \hat{\sigma}^2(\mu) \right) - \frac{n}{2} \quad . \quad (12) \\ \Rightarrow \frac{\partial \ln(L)}{\partial \mu} &= -\frac{n}{2} \frac{1}{\hat{\sigma}^2(\mu)} \frac{\partial \hat{\sigma}^2(\mu)}{\partial \mu} = 0\end{aligned}$$

- Given

$$\begin{aligned}\frac{\partial \bar{\sigma}^2(\mu)}{\partial \mu} &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu) = 0 \\ \Rightarrow \sum_{i=1}^n X_i - n\mu &= 0 \Rightarrow \hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^n X_i\end{aligned} \quad (13)$$

- Note that this solution is the same as

$$\begin{aligned}\ln(L) &= -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \\ \Rightarrow \frac{\partial \ln(L)}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu) \Rightarrow \hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^n X_i\end{aligned} \quad (14)$$

Normal Equations

- If we extend the above discussion to multiple regression, we can derive the normal equations.
 - Specifically, if

$$y_i = \alpha_0 + \alpha_1 x_i + \epsilon_i \quad (15)$$

where ϵ_i is distributed independently and identically normal, the concentrated likelihood function above can be rewritten as

$$\begin{aligned} \ln(L) &= -\frac{n}{2} \ln \left(\frac{1}{n} \sum_{i=1}^n [y_i - \alpha_0 - \alpha_1 x_i]^2 \right) \\ &= -\frac{n}{2} \ln \left(\frac{1}{n} \sum_{i=1}^n [y_i^2 - 2\alpha_0 y_i - 2\alpha_1 x_i y_i + \alpha_0^2 + 2\alpha_0 \alpha_1 x_i + \alpha_1^2 x_i^2] \right) \end{aligned} \quad (16)$$

- Taking the derivative with respect to α_0 yields

$$\begin{aligned} -\frac{n}{2} \frac{n}{\sum_{j=1}^n \left[(y_i - \alpha_0 - \alpha_1 x_i)^2 \right]} \sum_{i=1}^n [-2y_i + 2\alpha_0 + 2\alpha_1 x_i] &= 0 \\ \Rightarrow -\frac{1}{n} \sum_{i=1}^n y_i + \alpha_0 + \frac{\alpha_1}{n} \sum_{i=1}^n x_i &= 0 \\ \Rightarrow \alpha_0 = \frac{1}{n} \sum_{i=1}^n y_i - \alpha_1 \frac{\alpha_1}{n} \sum_{i=1}^n x_i & \end{aligned} \tag{17}$$

- Taking the derivative with respect to α_1 yields

$$\begin{aligned} -\frac{n}{2} \frac{n}{\sum_{j=1}^n [y_i - \alpha_0 - \alpha_1 x_i]^2} \sum_{i=1}^n [-2x_i y_i + 2\alpha_0 x_i + 2\alpha_1 x_i^2] &= 0 \\ \Rightarrow -\frac{1}{n} \sum_{i=1}^n x_i y_i + \frac{1}{n} \alpha_0 \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n \alpha_1 x_i^2 & \end{aligned} \quad (18)$$

- Substituting for α_0 yields

$$\begin{aligned} & -\frac{1}{n} \sum_{i=1}^n x_i y_i + \left(\frac{1}{n} \sum_{i=1}^n y_i \right) \left(\frac{1}{n} \sum_{i=1}^n x_i \right) + \\ & \alpha_1 \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \left(\frac{1}{n} \sum_{i=1}^n x_i \right) + \alpha_1 \frac{1}{n} \sum_{i=1}^n x_i^2 = 0 \end{aligned} \quad (19)$$

Hence

$$\alpha_1 = \frac{\frac{1}{n} \left(\sum_{i=1}^n x_i y_i - \left[\sum_{i=1}^n x_i \right] \left[\sum_{i=1}^n y_i \right] \right)}{\frac{1}{n} \left(\sum_{i=1}^n x_i^2 - \left[\sum_{i=1}^n x_i \right]^2 \right)} \quad (20)$$

Matrix Normal Equations

- The equations derived above can be written in a matrix form

$$\begin{pmatrix} \frac{1}{N} \sum_{i=1}^N y_i \\ \frac{1}{n} \sum_{i=1}^N x_i y_i \end{pmatrix} = \begin{bmatrix} 1 & \frac{1}{N} \sum_{i=1}^N x_i \\ \frac{1}{n} \sum_{i=1}^N x_i & \frac{1}{N} \sum_{i=1}^N x_i^2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \quad (21)$$

Properties of Maximum Likelihood Estimators

- Cramer-Rao Lower Bound

- **Theorem 7.4.1:** Let $L(X_1, X_2, \dots, X_n | \theta)$ be the likelihood function and let $\hat{\theta}(X_1, X_2, \dots, X_n)$ be an unbiased estimator of θ . Then, under general conditions, we have

$$V(\hat{\theta}) \geq -\frac{1}{E\left[\frac{\partial^2 \ln(L)}{\partial \theta^2}\right]} \quad (22)$$

- The right-hand side is known as the Cramer-Rao lower bound (CRLB).

Consistency of Maximum Likelihood

- The consistency of maximum likelihood can be shown by applying Khinchine's Law of Large Numbers to

$$Q_n(\theta) = \frac{1}{n} \ln(L_n(\theta)) = \frac{1}{n} \sum_{i=1}^n \ln(f[X_i|\theta]) \quad (23)$$

which converges as long as

$$E[\ln(f[X_i, \theta])] < \infty \quad (24)$$

Asymptotic Normality

- **Theorem 7.4.3:** Let the likelihood function be $L(X_1, X_2, \dots, X_n | \theta)$. Then, under general conditions, the maximum likelihood estimator of θ is asymptotically distributed as

$$\hat{\theta} \overset{A}{\sim} N \left(\theta, -E \left[\frac{\partial^2 \ln(L)}{\partial \theta^2} \right]^{-1} \right) \quad (25)$$

Readings

- Moss 8.1