ST 563 - Spring 2026
Week 7

This Week's Outline

Reading: *Statistical Inference 2nd Edition* Sections 9.1 - 9.3 (interval estimation)

Topics:

1. Confidence Intervals for Functions of Parameters
2. Bayes recap (Prior, Likelihood, Posterior)
3. Credible Probability
4. Credible Interval
5. Unimodal pdf
6. Highest Posterior Density Interval
7. Joint Confidence Intervals for Multivariate Parameters
8. Two-sample Confidence Intervals
9. Multivariate normal distribution

Results (Theorems/Lemmas/Propositions):

1. Theorem: Form of shortest interval containing given probability in unimodal distributions

Confidence Intervals for Functions of Parameters

Suppose that based on data $\mathbf{X}$, we construct a $(1-\alpha)$ confidence interval $\mathcal{C}_\theta$ for a parameter θ. Now suppose that we actually want a confidence interval for $\tau = g(\theta)$ for some function $g(\cdot)$. How can we obtain a confidence interval for τ?

1. Invert a test of $H_0 : \tau = \tau_0$ vs. $H_1 : \tau \neq \tau_0$.
2. Create a pivot $U = h(\mathbf{X}, \tau)$ and construct a pivotal interval for τ.
3. *OR* Transform the interval $\mathcal{C}_\theta$ into an interval for τ.

The first two of these methods are exactly what we have already covered; we are just using a different parameterization of our distribution. Note that we have to be a bit careful because some functions $g(\theta)$ do not produce well-identified distributions; for instance, if $X_1, \ldots, X_n \sim_{iid} N(\theta, 1)$ but we are interested in $\tau = \theta^2$, there are two values of θ for each value of τ, so knowing τ does not uniquely identify the distribution.

We can still perform inference for τ, we just have to acknowledge that τ is not a *one-to-one* mapping of the parameter θ.

Method 3 above is clearly the easiest approach, given that we already have a confidence interval θ to work with. Let $\mathcal{C}_\theta = (L_\theta(\mathbf{X}), U_\theta(\mathbf{X}))$. Then an interval for τ can be obtained by taking

$$\mathcal{C}_\tau = \{\tau_0 | \tau_0 = g(\theta_0) \text{ for some } \theta_0 \in \mathcal{C}_\theta\}$$

Two questions to address here:

1. What is the coverage level of the interval $\mathcal{C}_\tau$ constructed in this manner?

2. Why do we have to frame the interval like this? Why can't we instead write

$$\text{(WRONG:) } \mathcal{C}^*_\tau = (g(L_\theta(\mathbf{X})), g(U_\theta(\mathbf{X})))?$$

Bayesian Interval Estimators

Recall the basic components of Bayesian inference:

Definition 1. (Prior Distribution) The distribution $h(\theta)$ is called the prior distribution of the parameter ϑ.

Definition 2. (Likelihood Function) Given some value of the parameter $\vartheta = \theta$, the joint pdf/pmf of the data $\mathbf{x}$ is $f(\mathbf{x};\theta)$ is the likelihood (a function of both the value θ and the data $\mathbf{x}$).

Definition 3. (Posterior Distribution) The posterior distribution of ϑ given the observed data $\mathbf{x}$ is

$$k(\theta|\mathbf{x}) = \frac{f(\mathbf{x};\theta)h(\theta)}{\int_\theta f(\mathbf{x};\theta)h(\theta)d\theta}$$

Just like we used the posterior distribution to construct point estimates and decide between competing hypotheses, we can also use the posterior distribution to construct interval estimates.

First, we define the *credible probability* of any set $\mathcal{A} \subset \Theta$—that is, any subset of our parameter space:

Definition 4. For any region $\mathcal{A} \subset \Theta$, the credible probability of the set $\mathcal{A}$ is

$$P(\theta \in \mathcal{A}|\mathbf{x}) = \int_{\mathcal{A}} k(\theta|\mathbf{x})d\theta$$

Definition 5. A $1-\alpha$ *credible interval* is an interval $[a, b]$ such that the credible probability of that interval is $1-\alpha$; that is,

$$\int_a^b k(\theta|\mathbf{x})d\theta = 1-\alpha$$

Recall the Beta-Binomial Conjugate Family:

Example 1. (Beta-Binomial Conjugate Family) Suppose that $Y \sim \text{Binomial}(n, p)$, where p is unknown, and that we place a $\text{Beta}(\alpha, \beta)$ prior on p:

- The likelihood function is given by

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y} \text{ for } y = 0, \ldots, n$$

- and the prior is given by

$$h(p|\alpha,\beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} \qquad \text{for } p \in (0,1).$$

As we've seen, the posterior distribution density for p given our observed value $Y = y$ is

$$\begin{aligned} k(p|y) &= \frac{\binom{n}{y} p^y (1-p)^{n-y} B(\alpha,\beta)^{-1} p^{\alpha-1}(1-p)^{\beta-1}}{m(y)} \\ &\propto p^{y+\alpha-1}(1-p)^{n-y+\beta-1} \end{aligned}$$

which we recognize as the kernel of a Beta$(y+\alpha, n-y+\beta)$ distribution.

How would you construct a set $\mathcal{A}$ that has credible probability 0.95 (or more generally, $1-\alpha$ for any $\alpha \in [0,1]$)?

The following theorem tells us how to choose the *shortest* possible interval in cases where the posterior distribution is *unimodal*.

Definition 6. A pdf is *unimodal* if there exists some value x^* such that

- For all $x \le x^*$, $f(x)$ is non-decreasing
- For all $x \ge x^*$, $f(x)$ is non-increasing

Theorem 1. *Let $f(x)$ be a unimodal density function, and suppose a and b satisfy:*

1. $a \le x^* \le b$
2. $f(a) = f(b)$
3. $\int_a^b f(x)dx = 1-\alpha$

Then the interval $[a,b]$ is the ***shortest*** *interval among all intervals $[c,d]$ that satisfy $\int_c^d f(x)dx = 1-\alpha$.*

Proof.

- $\int_\theta^b k(\theta|\mathbf{x})d\theta = 1 - \alpha$

□

We can use the previous theorem to derive the shortest credible interval for a parameter given that its posterior distribution is unimodal:

Definition 7. For a unimodal posterior distribution $k(\theta|\mathbf{x})$, the highest posterior density $1 - \alpha$ credible interval is the interval $[a, b]$ such that

- $\int_\theta^b k(\theta|\mathbf{x})d\theta = 1 - \alpha$

- $k(a|\mathbf{x}) = k(b|\mathbf{x})$
- $a < \theta^* < b$ where θ^* is the mode of the posterior distribution.

The highest posterior density (HPD) $1-\alpha$ credible interval is, by the above theorem, the shortest $1-\alpha$ credible interval for θ. Note, however, that this interval might not contain the Bayes estimator for θ if the posterior distribution is very highly skewed and has a heavy tail.

Note that if the posterior is *not* unimodal, then the HPD region may consist of multiple disconnected intervals.

Example 2. (Poisson-Gamma conjugate family) Let $X_1, \ldots, X_n$ be iid Poisson(λ), and let λ have a Gamma(α, β) distribution, the conjugate family for the Poisson likelihood.

Derive the HDP $1-\alpha$ credible interval for λ when $a = b = 1$.

Joint Confidence Intervals for Multivariate Parameters

Suppose we have data $\mathbf{X}$ that provides information about a multivariate parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_p) \in \Theta \subset \mathcal{R}^p$, and we want to

1. Construct a confidence region $\mathcal{C}$ for θ
   - Note that this region $\mathcal{C} \subset \Theta \subset \mathcal{R}^p$ will be a p-dimensional region.
2. Perform a hypothesis test comparing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$.

This setting could arise in several ways:

1. Our data are iid from some distribution that has multiple parameters of interest; for instance,
   - Gamma(k, λ): $\theta = (k, \lambda)$
   - Normal(μ, σ^2): $\theta = (\mu, \sigma^2)$
   - Beta(α, β): $\theta = (\alpha, \beta)$
2. Our data come from several (possibly) distinct populations, and we want to construct a joint interval for the parameters of each of these populations; for instance,
   - $X_1, \ldots, X_n \sim$ iid Exponential(λ_X) and $Y_1, \ldots, Y_m \sim$ iid Exponential(λ_Y); $\theta = (\lambda_X, \lambda_Y)$.
   - Here our combined sample is $\mathbf{X} = \{X_1, \ldots, X_n, Y_1, \ldots, Y_m\}$.
3. Our data come from a multivariate distribution, which has multivariate parameters.
   - By far the most common example of this is the multivariate normal distribution, which we will discuss in more detail later. In this case, we might have $\theta = (\mu_1, \ldots, \mu_p)$, the mean vector.

We will see some special approaches for dealing with the third setting, multivariate normal data. For general multivariate parameters, however, we typically use the following approach for confidence regions:

1. Construct $(1 - \alpha^*)$ CIs $\mathcal{C}_j$ for each component θ_j of θ separately.
2. Define $\mathcal{C} = \{\theta_0 \mid \theta_{0,j} \in \mathcal{C}_j$ **for all** $j = 1, \ldots, p\}$, where $\theta_0 = (\theta_{0,1}, \theta_{0,2}, \ldots, \theta_{0,p})$.

   That is, our joint confidence region is the **intersection of all of the univariate confidence intervals** for each separate component of the parameter vector.
3. Chose the univariate coverage levels $(1 - \alpha^*)$ to ensure that the coverage level of the joint confidence region is *at least* the desired level $(1 - \alpha)$.

How do we accomplish the last step?

What if we know that the intervals that we are constructing are independent of each other? That is, the limits of $\mathcal{C}_j$ are independent of the limits of $\mathcal{C}_k$ for all $j \neq k$. Can we improve (shrink) the univariate confidence intervals while maintaining joint coverage probability of $(1 - \alpha)$?

Joint Confidence Interval for Normal Mean and Variance

Example 3. Let $X_1, \ldots, X_n$ be iid Normal(μ, σ^2) with both μ and σ^2 unknown. Find a joint $1 - \alpha$ confidence region for $\theta = (\mu, \sigma^2)$.

Note that the Bonferroni joint interval tends to *over-cover*; that is, we have

$$P\left((\mu, \sigma^2) \in \mathcal{C}_\mu \times \mathcal{C}_\sigma^2\right) = P(\mu \in \mathcal{C}_\mu, \sigma^2 \in \mathcal{C}_{\sigma^2}) > 1 - \alpha$$

We can work out the coverage probability under several different conditions:

1. **Perfect positive dependence** between the coverages of the two intervals:

   $$\mu \in \mathcal{C}_\mu \Leftrightarrow \sigma^2 \in \mathcal{C}_{\sigma^2}$$

   In this case,

   $$P(\mu \in \mathcal{C}_\mu, \sigma^2 \in \mathcal{C}_{\sigma^2}) = P(\mu \in \mathcal{C}_\mu) = P(\sigma^2 \in \mathcal{C}_{\sigma^2})$$

   so the two marginal intervals need only cover with probability $1 - \alpha$ to obtain a joint $1 - \alpha$ interval. This situation does not happen except in cases of redundant parameterization, but is worth considering as an extreme example of the result of positive dependence between the coverages.

2. **Independence** between the coverages of the two intervals:

$$\mu \in \mathcal{C}_\mu \perp\!\!\!\perp \sigma^2 \in \mathcal{C}_{\sigma^2}$$

In this case,

$$P(\mu \in \mathcal{C}_\mu, \sigma^2 \in \mathcal{C}_{\sigma^2}) = P(\mu \in \mathcal{C}_\mu) \times P(\sigma^2 \in \mathcal{C}_{\sigma^2})$$

so the two marginal intervals need to cover with probability $1-\alpha_1$ and $1-\alpha_2$ where $(1-\alpha_1)(1-\alpha_2) = (1-\alpha)$ to obtain a joint $1-\alpha$ interval (e.g. $(1-\alpha_1) = (1-\alpha_2) = (1-\alpha)^{1/2}$).

3. **Perfect (largest possible) negative dependence** between the coverages of the two intervals:

$$\mu \notin \mathcal{C}_\mu \Rightarrow \sigma^2 \in \mathcal{C}_{\sigma^2}$$
$$\sigma^2 \notin \mathcal{C}_{\sigma^2} \Rightarrow \mu \in \mathcal{C}_\mu$$

That is, the events $\mu \notin \mathcal{C}_\mu$ and $\sigma^2 \notin \mathcal{C}_\sigma^2$ are mutually exclusive. In this case,

$$\begin{aligned} P(\mu \in \mathcal{C}_\mu, \sigma^2 \in \mathcal{C}_{\sigma^2}) &= 1 - P(\mu \notin \mathcal{C}_\mu \cup \sigma^2 \notin \mathcal{C}_{\sigma^2}) \\ &= 1 - \left[P(\mu \notin \mathcal{C}_\mu) + P(\sigma^2 \notin \mathcal{C}_{\sigma^2})\right] \end{aligned}$$

so the two marginal intervals need to cover with probabilities $1-\alpha_1$ and $1-\alpha_2$ where $\alpha_1 + \alpha_2 = \alpha$ to obtain a joint $1-\alpha$ interval. (e.g. $\alpha_1 = \alpha_1 = \alpha/2$.)

Two-sample Confidence Intervals

In a two sample problem, we have data $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from two possibly different populations/distributions, and we would like to make inference regarding some comparison of parameters for the distributions.

- $X_1, \ldots, X_m$ iid Exp(λ_X) $Y_1 \ldots, Y_n$ iid Exp(λ_Y)
  - Parameter $\theta = (\lambda_X, \lambda_Y)$
  - Comparison of interest might be $\kappa(\theta) = \lambda_X/\lambda_Y$.
- $X_1, \ldots, X_m$ iid Uniform$(0, \tau_X)$ $Y_1 \ldots, Y_n$ iid Uniform$(0, \tau_Y)$
  - Parameter $\theta = (\tau_X, \tau_Y)$
  - Comparison of interest might be $\kappa(\theta) = \tau_X/\tau_Y$.
- $X_1, \ldots, X_m$ iid Normal(μ_X, σ_X^2) $Y_1, \ldots, Y_n$ iid Normal(μ_Y, σ_Y^2)
  - Parameter $\theta = (\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2)$
  - Comparison of interest might be $\kappa(\theta) = \mu_X - \mu_Y$ or $\kappa(\theta) = \sigma_X^2/\sigma_Y^2$.

In general, the same approaches that were used in the one-sample setting can be used for constructing confidence intervals for these comparisons of parameters:

1. Invert a hypothesis test for $H_0 : \kappa(\theta) = \kappa_0$ to obtain
   - One-sided upper confidence interval

     $$\mathcal{C} = (-\infty, T_U(\mathbf{X}, \mathbf{Y}))$$

     if the alternative is a one-sided lower hypothesis $H_1 : \kappa(\theta) < \kappa_0$.

- One-sided lower confidence interval

$$\mathcal{C} = (T_L(\mathbf{X}, \mathbf{Y}), \infty)$$

  if the alternative is a one-sided upper hypothesis $H_1 : \kappa(\theta) > \kappa_0$.
- Two-sided upper confidence interval

$$\mathcal{C} = (T_L(\mathbf{X}, \mathbf{Y}), T_U(\mathbf{X}, \mathbf{Y}))$$

  if the alternative is a two-sided hypothesis $H_1 : \kappa \neq \kappa_0$.

However, we have not yet seen hypothesis test procedures for the two-sample setting–we will get to these in the next few weeks. That leaves us Option 2:

2. Pivotal method:

   (a) Construct a pivot

$$U = h(\mathbf{X}, \mathbf{Y}, \theta)$$

   such that the distribution of U does not depend on the value of the comparison parameter $\kappa(\theta)$. Let

$$F_U(u) = P(U \leq u)$$

   be the cdf of the pivot.

   (b) Find a and/or b such that

$$\begin{aligned} P(a_1 < U) &= 1 - \alpha && \text{For a one-sided upper CI for } \theta \\ P(U \leq b_1) &= 1 - \alpha && \text{For a one-sided lower CI for } \theta \\ P(a_2 < U \leq b_2) &= 1 - \alpha && \text{For a two-sided CI for } \theta \end{aligned}$$

   This is done by letting

$$\begin{aligned} a_1 &= F_U^{-1}(\alpha) \\ b_1 &= F_U^{-1}(1 - \alpha) \\ \\ a_2 &= F_U^{-1}(\alpha/2) \\ b_2 &= F_U^{-1}(\alpha/2) \end{aligned}$$

   (c) Solve the inequalities

$$\begin{aligned} a_1 &< U \\ U &\leq b_1 \\ a_2 < &U \leq b_2 \end{aligned}$$

   for $\kappa(\theta)$.

The general form for pivots for location and scale parameter families is similar in the two-sample case to the one-sample case.

- Location parameter:

- Point estimator $\widehat{\kappa(\theta)}$ for $\kappa(\theta)$ based on sufficient statistic $\mathbf{T}$ for θ
  - Often the distribution of $U = \widehat{\kappa(\theta)} - \kappa(\theta)$ will not depend on θ.
  - Note that this is instead presented as

$$U = \frac{\widehat{\kappa(\theta)} - \kappa(\theta)}{\tau}$$

    for some "nuisance" parameter τ that is either known or estimated from the data. This is just to get the pivot distribution to be of a more standard form (e.g. Standard normal instead of Normal$(0, \sigma^2)$).
  - When τ is estimated from the data, we need to account for the randomness introduced by this estimation in deriving the pivot distribution (e.g. when the normal variance is estimated instead of known, the pivot distribution is t_{n-1} instead of Normal(0, 1)).

- Scale parameter:
  - Point estimator $\widehat{\kappa(\theta)}$ for $\kappa(\theta)$ based on sufficient statistic $\mathbf{T}$ for θ
  - Often the distribution of $U = \widehat{\kappa(\theta)}/\kappa(\theta)$ will not depend on θ.

Example 4. (Exponential Location with Unknown Scale) Let

$$\begin{aligned} X_1, X_2, \ldots, X_n &\overset{\text{iid}}{\sim} \text{Exponential}(\mu_X, \sigma) \\ Y_1, Y_2, \ldots, Y_n &\overset{\text{iid}}{\sim} \text{Exponential}(\mu_Y, \sigma) \end{aligned}$$

so

$$\begin{aligned} f(x; \mu_X, \sigma) &= \frac{1}{\sigma} e^{-(x-\mu_X)/\sigma} \qquad \text{for } x > \mu_X \\ f(y; \mu_Y, \sigma) &= \frac{1}{\sigma} e^{-(y-\mu_Y)/\sigma} \qquad \text{for } x > \mu_Y \end{aligned}$$

Find a $(1-\alpha)$ confidence interval for $\kappa(\theta) = \mu_X - \mu_Y$ if σ is unknown (but the same value for both populations).

Multivariate Normal Distribution

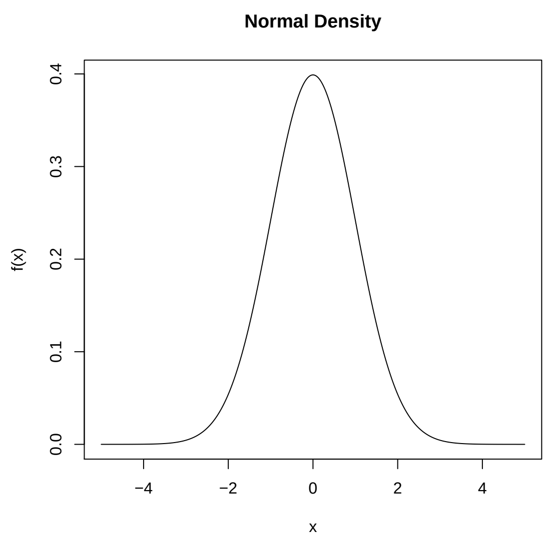
Recall the univariate normal (also called Gaussian) distribution with parameters μ (expectation) and σ^2 (variance) has density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for } -\infty < x < \infty$$

The cumulative distribution function

$$F(x) = P(X \le x) = \int_{u=-\infty}^{x} f(u)\, du$$

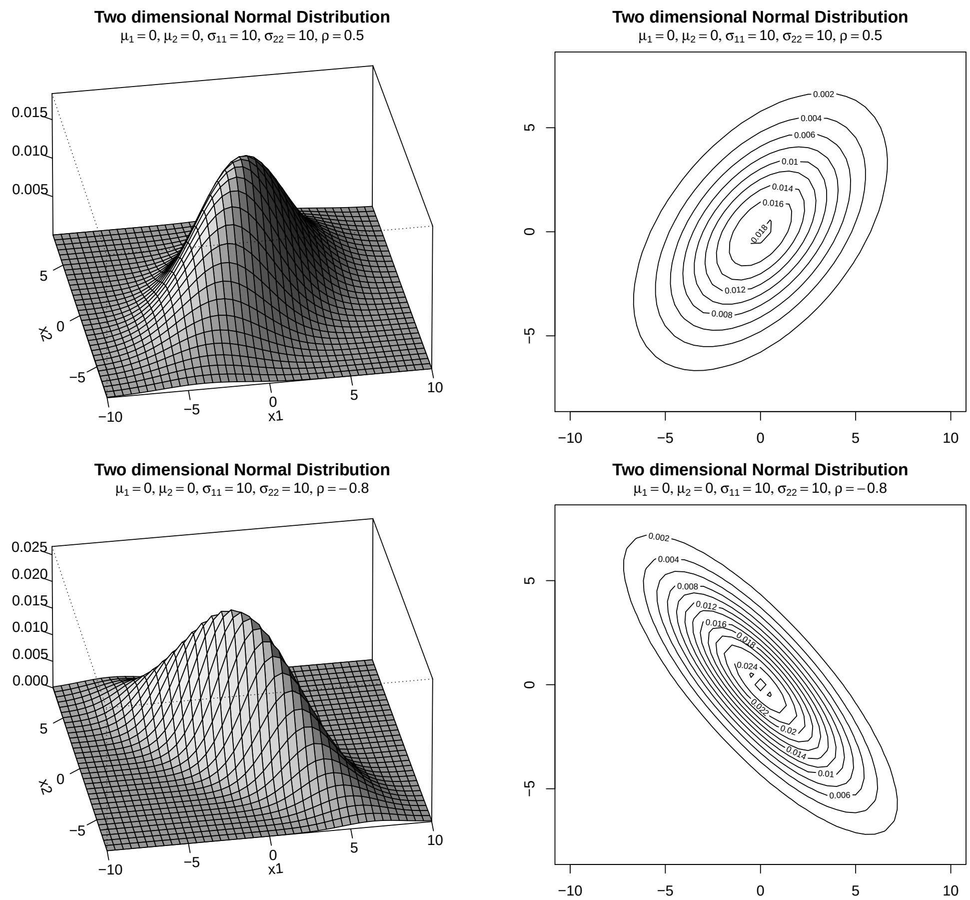
does not have a closed form expression.



The multivariate normal distribution is an extension of the univariate normal distribution. A random vector $\mathbf{X} = [X_1, X_2, \ldots, X_p]^T$ has a multivariate normal distribution with parameters μ (expectation vector) and $\mathbf{\Sigma}$ (covariance matrix) if the density for the joint distribution of the elements of $\mathbf{X}$ is

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_p) = \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}$$

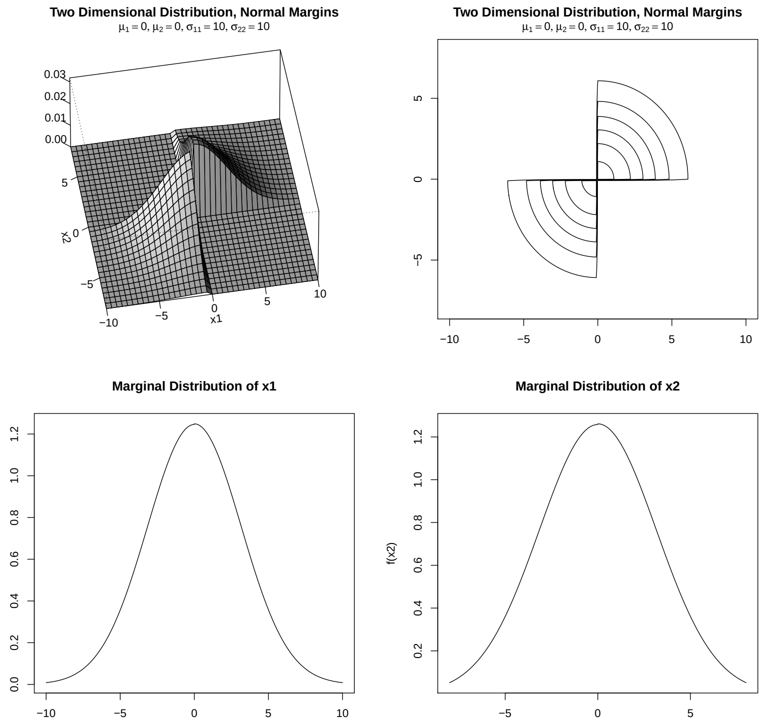
Clearly the density will have constant contours given by



Properties of the Multivariate Normal Distribution

- If $\mathbf{X} = [X_1, X_2, \ldots, X_p]^T$ has a multivariate normal distribution, then each element $X_j, j = 1, \ldots, p$ has a *marginal* normal distribution; that is, each element considered on its own is normally distributed with mean μ_j and variance σ_{jj}.
- Random variables $X_1, X_2, \ldots, X_p$ with *marginal* normal distributions do NOT necessarily have a multivariate normal joint distribution.

Example 5. (Normal margins but not multivariate normal joint distribution)



Properties of the Multivariate Normal Distribution

- All subsets of elements of $\mathbf{X}$ have a multivariate normal distribution.

- If $\mathbf{X} = [X_1, X_2, \ldots, X_p]^T$ has a multivariate normal distribution, then all linear combinations of the components of $\mathbf{X}$ are normally distributed.

- If $\mathbf{X} = [X_1, X_2, \ldots, X_p]^T$ has a multivariate normal distribution MVN($\boldsymbol{\mu}, \boldsymbol{\Sigma}$) and $\mathbf{c} = [c_1, c_2, \ldots, c_p]^T$ is a vector of constants, then

- If $\mathbf{X} = [X_1, X_2, \ldots, X_p]^T$ has a multivariate normal distribution, then for a constant matrix $\mathbf{A}_{(q\times p)}$, the $(q \times 1)$ vector

$$\mathbf{AX} = \begin{bmatrix} a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \\ a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + a_{q2}X_2 + \ldots + a_{qp}X_p \end{bmatrix}$$

has a multivariate normal distribution with mean $\mathbf{A}\boldsymbol{\mu}$ and covariance $\mathbf{A\Sigma A}^T$. That is,

$$\mathbf{AX} \sim \text{MVN}\left(\mathbf{A}\mu, \mathbf{A\Sigma A}^T\right)$$

- If $\mathbf{X} = [X_1, X_2, \ldots, X_p]^T$ has a multivariate normal distribution, then $\text{Cov}(X_j, X_k) = \sigma_{jk} = 0$ if and only if X_j and X_k are independent.

  This is NOT true for general random variables X_j and X_k: it is possible for X_j and X_k to be highly dependent (i.e., knowing the value of X_j provides a lot of information about the value of X_k) but have $\text{Cov}(X_j, X_k) = 0$.

Example 6. To illustrate this last point, let

$$X \sim \text{Normal}(0,1) \qquad Y \sim \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$$

with Y independent of X. Define

$$W = Y \times X \qquad \text{so} \qquad W = \begin{cases} X & \text{w.p. } \frac{1}{2} \\ -X & \text{w.p. } \frac{1}{2} \end{cases}$$

1. What is the distribution of W?

2. What is the covariance between X and W?

3. Are X and W independent?

For a univariate normal random variable $X \sim \text{Normal}(\mu, \sigma^2)$, the *standardized* random variable

$$Z = \frac{X - \mu}{\sigma} \sim \text{Normal}(0, 1)$$

has a *standard normal distribution*; that is, a normal distribution with mean 0 and variance 1.

Using the properties of the multivariate normal distribution, we can show that a similar result holds for a multivariate normal random vector:

We have just seen that the elements $Z_1, Z_2, \ldots, Z_p$ of the standardized random vector

$$\underset{(p\times 1)}{\mathbf{Z}} = \mathbf{\Sigma}^{-1/2}(\mathbf{X} - \mu)$$

are independent, identically distributed (iid) random variables, each with a standard normal distribution.

Recall the following fact: for $Y_1, Y_2, \ldots, Y_k \overset{iid}{\sim} \text{Normal}(0, 1)$, we have

$$Y_1^2 + Y_2^2 + \ldots + Y_k^2 = \sum_{i=1}^{k} Y_i^2 \sim \chi^2_{(k)};$$

that is, the sum of the squared random variables $Y_1, \ldots, Y_k$ has a chi-squared distribution with k degrees of freedom.

Therefore, the sum of the squared elements of the standardized vector $\mathbf{Z}$ have a chi-squared distribution with p degrees of freedom:

$$Z_1^2 + Z_2^2 + \ldots + Z_p^2 = \sum_{i=1}^{p} Z_i^2 = \mathbf{Z}^T\mathbf{Z} \sim \chi^2_{(p)}$$

Note that

$$\begin{aligned}\mathbf{Z}^T\mathbf{Z} &= \left[\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\mu)\right]^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\mu) \\ &= (\mathbf{X}-\mu)^T \boldsymbol{\Sigma}^{-1/2^T}\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\mu) \\ &= (\mathbf{X}-\mu)^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mu)\end{aligned}$$

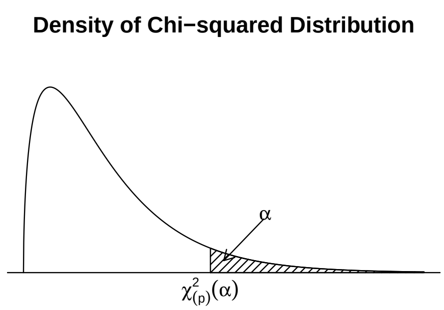
where we have used the following two facts:

- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
- $\boldsymbol{\Sigma}^{-1/2}$ is symmetric, so $\boldsymbol{\Sigma}^{-1/2^T} = \boldsymbol{\Sigma}^{-1/2}$

Using the fact that $\mathbf{Z}^T\mathbf{Z} \sim \chi^2_{(p)}$ and that $\mathbf{Z}^T\mathbf{Z} = (\mathbf{X}-\mu)^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mu)$, we obtain the following result:

$$P\left((\mathbf{X}-\mu)^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mu) < \chi^2_{(p)}(\alpha)\right) = 1-\alpha$$

where $\chi^2_{(p)}(\alpha)$ is the upper α quantile of the chi-squared distribution with p degrees of freedom:



Recall that $(\mathbf{x}-\mu)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu) = c^2$ defines a constant contour of the multivariate normal density; that is, for all vectors $\mathbf{x}$ such that

$$(\mathbf{x}-\mu)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu) = c^2$$

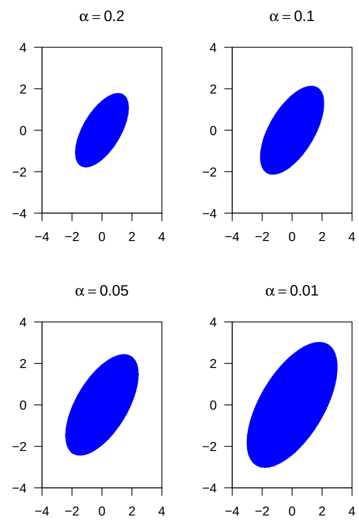
the density

$$f_{X_1,\ldots,X_p}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu)}$$

will have the same value. Using an earlier result, we now know that for $c^2 = \chi^2_{(p)}(\alpha)$, the probability

$$P\left((\mathbf{x}-\mu)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu) \le c^2\right) = 1-\alpha$$

That is, the area contained by the $\chi^2_{(p)}(\alpha)$ contour has probability $1-\alpha$. This result is useful (later) for constructing confidence intervals for the parameter vector μ.



So far, we've been dealing with a single random vector $\mathbf{X}$ (one observation).

Now we consider the case where we have a sample of n random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, **independent**, with

$$\underset{(p\times 1)}{\mathbf{X}_i} \sim \text{Normal}\left(\underset{(p\times 1)}{\mu_i}, \underset{(p\times p)}{\mathbf{\Sigma}}\right)$$

- The mean vectors μ_i may differ (for the sake of complete generality, though later we will consider the case where the $\mathbf{X}_i$ are iid)
- The covariance matrix $\mathbf{\Sigma}$ is assumed to be the same for all i.

This is analogous to the univariate setting where $X_1, \ldots, X_n$ are independent with $X_i \sim \text{Normal}(\mu_i, \sigma^2)$, where the means μ_i may differ, but σ^2 is assumed to be the same for all i.

In the univariate case, we have the following result:

$$V_1 = c_1X_1 + c_2X_2 + \ldots + c_nX_n = \sum_{i=1}^n c_iX_i \sim \text{Normal}\left(\sum_{i=1}^n c_i\mu_i, \left(\sum_{i=1}^n c_i^2\right)\sigma^2\right)$$

A completely analogous result holds for the multivariate case:

$$\mathbf{V}_1 = c_1\mathbf{X}_1 + c_2\mathbf{X}_2 + \ldots + c_n\mathbf{X}_n = \sum_{i=1}^{n} c_i\mathbf{X}_i \sim \text{Normal}\left(\sum_{i=1}^{n} c_i\boldsymbol{\mu}_i, \left(\sum_{i=1}^{n} c_i^2\right)\boldsymbol{\Sigma}\right)$$

We have just replaced the univariate random variables X_i with the multivariate random vectors $\mathbf{X}_i$, the univariate means μ_i with the vector means $\boldsymbol{\mu}_i$, and the univariate variance σ^2 with the multivariate covariance matrix $\boldsymbol{\Sigma}$.

A special, but by far most important and most common, case of the previous results occurs when the means $\mu_i = \mu$ (univariate) or $\boldsymbol{\mu}_i = \boldsymbol{\mu}$ (multivariate) are all equal, and the coefficients $\mathbf{c}_i = \frac{1}{n}$. We consider the univariate case first:

$$V_1 = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Then we have

$$\begin{aligned}\bar{X} &\sim \text{Normal}\left(\sum_{i=1}^{n}\frac{1}{n}\mu, \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2\sigma^2\right)\\ &= \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)\end{aligned}$$

The analogous result for the multivariate case:

$$\mathbf{V}_1 = \frac{1}{n}\mathbf{X}_1 + \frac{1}{n}\mathbf{X}_2 + \ldots + \frac{1}{n}\mathbf{X}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i = \bar{\mathbf{X}}$$

so we have that

Thus we have obtained the *sampling distribution* of the sample mean vector.

The sampling distribution for the sample covariance matrix

$$\underset{(p\times p)}{\mathbf{S}} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$$

may be similarly found to be analogous to the univariate case, which we now review.

In the univariate case, the sample variance is

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

The distribution of S^2 when $X_1, \ldots, X_n$ are iid Normal(μ, σ^2) is given by

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

That is, the ratio of $n-1$ times the sample variance divided by the true variance has a chi-squared distribution with $n-1$ degrees of freedom.

Recall, again, that a chi-squared distribution with k degrees of freedom is the distribution of

$$Z_1^2 + Z_2^2 + \ldots + Z_k^2 = \sum_{i=1}^{k} Z_i^2$$

for $Z_i \overset{iid}{\sim}$ Normal$(0, 1)$.

The multivariate version of this result states that

$$\underset{(p\times p)}{\mathbf{S}} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$$

has distribution given by

$$(n-1)\underset{(p\times p)}{\mathbf{S}} \sim \text{Wishart}\,(n-1, \boldsymbol{\Sigma})$$

or equivalently (as we will see shortly)

$$(n-1)\boldsymbol{\Sigma}^{-1/2}\mathbf{S}\boldsymbol{\Sigma}^{-1/2} \sim \text{Wishart}\,(n-1, \mathbf{I}_p)$$

The Wishart$(k, \boldsymbol{\Sigma})$ distribution is a distribution on $(p \times p)$ matrices, and may be obtained as the distribution of

$$\sum_{i=1}^{k} \underset{(p\times 1)}{\mathbf{Z}_i}\,\underset{(1\times p)}{\mathbf{Z}_i^T}$$

for random vectors

$$\mathbf{Z}_i \overset{iid}{\sim} \text{Normal}\left(\underset{(p\times 1)}{\mathbf{0}}, \underset{(p\times p)}{\boldsymbol{\Sigma}}\right)$$

This clearly reduces to the chi-square distribution in the univariate case (when $p = 1$) if $\boldsymbol{\Sigma} = \mathbf{I}_p = 1$.

Several properties of the chi-squared distribution extend to the Wishart distribution.

- Properties of chi-squared random variables:

- For $W_1 \sim \chi^2_{(\nu_1)}$ and $W_2 \sim \chi^2_{(\nu_2)}$, we have $W_1 + W_2 \sim \chi^2_{(\nu_1+\nu_2)}$.
- For $W_i \sim \chi^2_{(\nu_i)}$ for $i = 1, \ldots, m$, we have $\sum_{i=1}^m W_i \sim \chi^2_{(\nu)}$ where $\nu = \sum_{i=1}^m \nu_i$.

- Properties of Wishart random matrices:
  - For $\mathbf{W}_1 \sim \text{Wishart}(\nu_1, \mathbf{\Sigma})$ and $\mathbf{W}_2 \sim \text{Wishart}(\nu_2, \mathbf{\Sigma})$ we have $\mathbf{W}_1 + \mathbf{W}_2 \sim \text{Wishart}(\nu_1 + \nu_2, \mathbf{\Sigma})$.
  - For $\mathbf{W}_i \sim \text{Wishart}(\nu_i, \mathbf{\Sigma})$ for $i = 1, \ldots, m$, we have $\sum_{i=1}^m \mathbf{W}_i \sim \text{Wishart}(\nu, \mathbf{\Sigma})$ where $\nu = \sum_{i=1}^n \nu_i$.

Another property of the Wishart distribution: If $\mathbf{A}$ is distributed according to the Wishart$(\nu, \mathbf{\Sigma})$ distribution, then $\mathbf{CAC}^T$ is distributed according to the Wishart$(\nu, \mathbf{C\Sigma C}^T)$ distribution.

This result is the reason for the earlier equivalence:

$$(n-1)\underset{(p\times p)}{\mathbf{S}} \sim \text{Wishart}\,(n-1, \mathbf{\Sigma})$$

$$\Updownarrow$$

$$(n-1)\mathbf{\Sigma}^{-1/2}\mathbf{S}\mathbf{\Sigma}^{-1/2} \sim \text{Wishart}\,(n-1, \mathbf{I}_p)$$

Hypothesis tests for Multivariate Normal Parameters

Example 7. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, **independent**, with

$$\underset{(p\times 1)}{\mathbf{X}_i} \sim \text{Normal}\left(\underset{(p\times 1)}{\mu}, \underset{(p\times p)}{\mathbf{\Sigma}}\right)$$

If $\mathbf{\Sigma}$ is known, or if $\mathbf{\Sigma} = \sigma^2\mathbf{H}$ where H is known, we can use the sampling distribution of the sample mean vector to construct a confidence region for μ. This confidence region can be inverted to obtain a hypothesis test for $H_0 : \mu = \mu_{\mathbf{0}}$ vs. $H_1 : \mu \neq \mu_{\mathbf{0}}$.
