

MTH 511a - 2020: Lecture 10

Instructor: Dootika Vats

The instructor of this course owns the copyright of all the course materials. This lecture material was distributed only to the students attending the course MTH511a: “Statistical Simulation and Data Analysis” of IIT Kanpur, and should not be distributed in print or through electronic media without the consent of the instructor. Students can make their own copies of the course materials for their use.

1 Miscellaneous methods in sampling

1.1 Known relationships

It is always useful to remember the relationships between different distributions.

1. **Binomial distribution:** We know that if $Y_1, Y_2, \dots, Y_n \stackrel{iid}{\sim} \text{Bern}(p)$, then

$$X = Y_1 + Y_2 + \dots + Y_n \sim \text{Bin}(n, p).$$

So, we can simulate n Bernoulli variables, add them up, and we have a realization from a Binomial(n, p).

2. **Negative binomial distribution:** Number of failures until the r th success. So possibly related to geometric! If $Y_1, Y_2, \dots, Y_r \stackrel{iid}{\sim} \text{Geom}(p)$ (on failures), then

$$X = Y_1 + Y_2 + \dots + Y_r \sim \text{NB}(r, p).$$

3. **Beta distribution** If $X \sim \text{Gamma}(a, 1)$ and $Y \sim \text{Gamma}(b, 1)$, then

$$\frac{X}{X + Y} \sim \text{Beta}(a, b).$$

4. **Dirichlet distribution** : The Dirichlet distribution is a distribution over pmf.

$$f(x_1, x_2, \dots, x_k) = \frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i - 1} \quad 0 \leq x_i \leq 1, \sum_{i=1}^k x_i = 1.$$

The Dirichlet distribution is a generalization of the Beta distribution. Similarly,

$$\begin{aligned} Y_1 &\sim \text{Gamma}(\alpha_1, 1) \\ Y_2 &\sim \text{Gamma}(\alpha_2, 1) \\ &\vdots \\ Y_k &\sim \text{Gamma}(\alpha_k, 1) \end{aligned}$$

Let

$$X_i = \frac{Y_i}{\sum_{i=1}^k Y_i}.$$

Then $(X_1, \dots, X_k) \sim \text{Dir}(\alpha_1, \alpha_2, \dots, \alpha_k)$.

5. **Chi-squared distribution**: If $Y_1, Y_2, \dots, Y_k \stackrel{iid}{\sim} N(0, 1)$, then

$$X = Y_1^2 + Y_2^2 + \dots + Y_k^2 \sim \chi_k^2.$$

This way we can simulate χ^2 distributions with integer degrees of freedom.

6. **t-distribution** Let $Z \sim N(0, 1)$ and $Y \sim \chi_k^2$, then

$$X = \frac{Z}{\sqrt{\frac{Y}{k}}} \sim t_k.$$

7. **Location-scale family**: If Z has CDF $F_Z(z)$ in the sense that it doesn't have any parameters. Then for $\mu \in R$ and $\sigma > 0$,

$$Y = \mu + \sigma Z \text{ has CDF } F_X\left(\frac{z - \mu}{\sigma}\right).$$

If Z has pdf $f(z)$ then Y has pdf $\sigma^{-1}f((z - \mu)/\sigma)$.

So, if $Z \sim N(0, 1)$, then $Y = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

1.2 Sampling from mixture distributions

Mixture distributions for continuous densities is very similar to the discrete distributions. For $j = 1, \dots, k$, let $F_j(x)$ be a distribution function and let $f_j(x)$ be the corresponding density function. A mixture distribution function, F is

$$F(x) = \sum_{j=1}^k \pi_j F_j(x) \quad \text{where } \pi_j > 0, \sum_j \pi_j = 1$$

and the corresponding density is

$$f(x) = \sum_{j=1}^k \pi_j f_j(x).$$

If we can draw from each F_j then we can draw from the mixture distribution F as well using the same steps as in the discrete case.

Algorithm 1 Sampling from a mixture distribution

- 1: Generate $U \sim U[0, 1]$
 - 2: If $U < \pi_1$, generate from F_1
 - 3: If $U < \pi_1 + \pi_2$, generate from F_2
 - 4: ...
 - 5: If $U < \sum_{i=1}^{k-1} \pi_i$, generate from F_{k-1}
 - 6: Else, generate from F_k .
-

Example 1 (Mixture of normals). Consider two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. For some $0 < p < 1$, the mixture density is

$$\begin{aligned} f(x) &= p f_1(x; \mu_1, \sigma_1^2) + (1-p) f_2(x; \mu_2, \sigma_2^2) \\ &= p \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} + (1-p) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right\} \end{aligned}$$

Mixture distributions are particularly useful for clustering problems and we will come back to them again in the data analysis part of the course. If we want to sample from this distribution

Algorithm 2 Sampling from a Gaussian mixture

- 1: Generate $U \sim U[0, 1]$
 - 2: If $U < p$, generate $N(\mu_1, \sigma_1^2)$ (using location-scale family trick)
 - 3: Otherwise, generate $N(\mu_2, \sigma_2^2)$.
-

Example 2 (Zero-inflated gamma distribution). Just like the zero-inflated Poisson distribution, there are zero-inflated normal and Gamma distributions. Let's motivate the zero-inflated Gamma distribution:

Suppose you are an auto-insurance company and you want to study the cost of claims associated with each customer. That is, each customer, if they have an accident, will come to you and claim insurance money reimbursement for the accident. So

Let X = insurance money asked for by a customer in a month.

However, most customers will not enter into any accidents, so they will claim Rs 0. But when they do, they will claim reimbursement for some amount of money that, say, will follow a Gamma distribution.

The density function can be defined as follows for $0 < p < 1$

$$f(x) = p\mathbb{I}(x = 0) + (1 - p)\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-x\beta}.$$

Algorithm 3 Sampling from a zero-inflated Gamma

- 1: Generate $U \sim U[0, 1]$
 - 2: If $U < p$, return $X = 0$
 - 3: Otherwise, generate $X \sim \text{Gamma}(\alpha, \beta)$.
-

1.3 Multidimensional target

We have almost entirely focused on univariate densities, but most often interest is in multivariate/multidimensional target distribution.

- **Conditional Distribution:** Consider a variable $\mathbf{X} = (X_1, X_2, \dots, X_k)$, with a joint pdf

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_k).$$

We can use conditional distribution properties:

$$f(\mathbf{x}) = f_{X_1}(x_1)f_{X_2|X_1}(x_2) \cdots f_{X_k|X_1, \dots, X_{k-1}}(x_k).$$

Algorithm 4 Sampling \mathbf{X} using conditional distributions

- 1: Generate $X_1 \sim f_{X_1}(x_1)$
 - 2: Generate $X_2 \sim f_{X_2|X_1}(x_2)$
 - 3: Generate $X_3 \sim f_{X_3|X_2, X_1}(x_3)$
 - 4: \vdots
 - 5: Generate $X_n \sim f_{X_n|X_{k-1}, \dots, X_1}(x_k)$
 - 6: Return $\mathbf{X} = (X_1, \dots, X_k)$
-

- **Multivariate normal:** Consider sampling from a $N_k(\mu, \Sigma)$ where Σ is positive definite. Then for $|\cdot|$ denoting determinant,

$$f_{\mathbf{x}}(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{k/2} |\Sigma|^{-1/2} \exp\left\{-\frac{(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)}{2}\right\},$$

is the density of a multivariate normal distribution with mean μ and covariance Σ . First, note that since Σ is a positive-definite (symmetric) matrix, we can use the spectral decomposition

$$\Sigma = Q\Lambda Q^{-1}$$

where Q is the matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues. Then, we can define the *square-root* of Σ as

$$\Sigma^{1/2} = Q\Lambda^{1/2}Q^{-1},$$

so that

$$\Sigma^{1/2}\Sigma^{1/2} = Q\Lambda^{1/2}Q^{-1}Q\Lambda^{1/2}Q^{-1} = Q\Lambda Q^{-1}.$$

Similarly, the inverse square-root is

$$\Sigma^{-1/2} = Q\Lambda^{-1/2}Q^{-1},$$

Set $\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \mu)$. Then

$$\mathbf{Z} \sim N_k(0, I_k).$$

That is, \mathbf{Z} is a k -dimensional multivariate normal distribution with an identity covariance matrix. Which implies if $\mathbf{Z} = (Z_1, \dots, Z_k)$, then $\text{Cov}(Z_i, Z_j) = 0$ for all $i \neq j$.

For the normal distribution, if the covariance is zero, then the random variables are independent! This isn't true in general but is true for normal random variables.

So, to sampling from $N_k(\mu, \Sigma)$, we can sample $Z_1, Z_2, \dots, Z_k \stackrel{iid}{\sim} N(0, 1)$, and set $\mathbf{Z} = (Z_1, \dots, Z_k)$. Then

$$\mathbf{X} := \mu + \Sigma^{1/2}\mathbf{Z} \sim N_k(\mu, \Sigma).$$

Then $\mathbf{Z} \sim N_k(\mu, \Sigma)$.

```
#####
## Generate from multivariate normal distribution
#####
set.seed(1)
par(mfrow = c(2,2))

# Function produces samples from a multivariate normal
multinorm <- function(mu, Sigma, N = 5e2)
{
  # Eigenvalue (spectral) decomposition
  decomp <- eigen(Sigma)

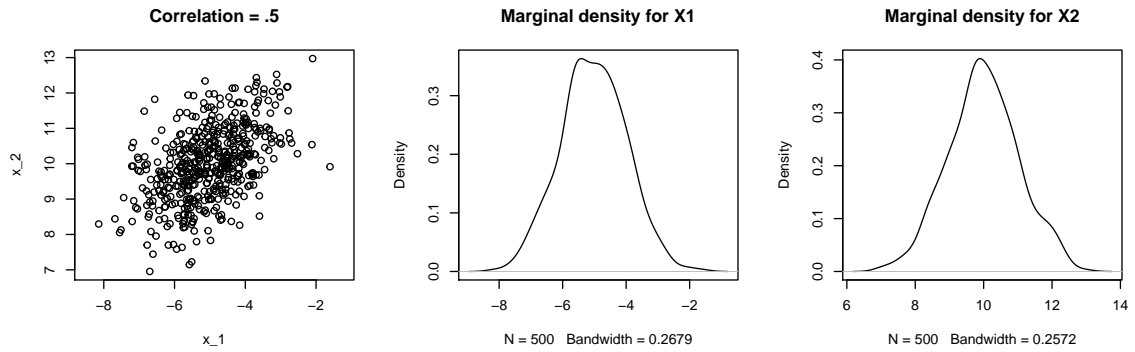
  # Finding matrix square-root
  Sig.sq <- decomp$vectors %*% diag(decomp$values^(1/2)) %*%
    solve(decomp$vectors)

  samp <- matrix(0, nrow = N, ncol = 2)
  for(i in 1:N)
  {
    Z <- rnorm(2)
    samp[i, ] <- mu + Sig.sq %*%Z
  }
  return(samp)
}

###
# First: Mean (-5, 10) and .5 correlation
mu <- c(-5,10)
Sigma <- matrix(c(1, .5, .5, 1), nrow = 2, ncol = 2)
samp <- multinorm(mu = mu, Sigma = Sigma)

par(mfrow = c(1,3))
plot(samp, asp = 1, main = "Correlation = .5", xlab = "x_1", ylab =
  "x_2")
plot(density(samp[,1]), main = "Marginal density for X1")
plot(density(samp[,2]), main = "Marginal density for X2")

par(mfrow = c(2,2))
plot(samp, asp = 1, main = "Correlation = .5", xlab = "x_1", ylab =
  "x_2")
```



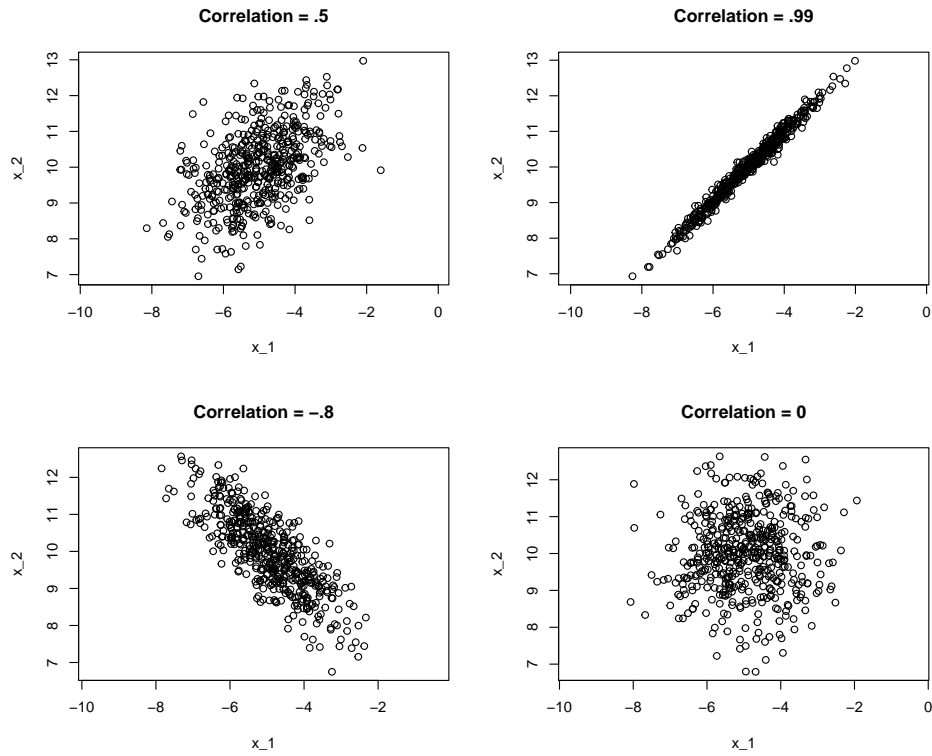
```

###
# Second: Mean (-5, 10) and .99 correlation
mu <- c(-5,10)
Sigma <- matrix(c(1, .99, .99, 1), nrow = 2, ncol = 2)
plot(samp, asp = 1, main = "Correlation = .99", xlab = "x_1", ylab =
     "x_2")

###
# Third: Mean (-5, 10) and -.8 correlation
mu <- c(-5,10)
Sigma <- matrix(c(1, -.8, -.8, 1), nrow = 2, ncol = 2)
samp <- multinorm(mu = mu, Sigma = Sigma)
plot(samp, asp = 1, main = "Correlation = -.8", xlab = "x_1", ylab =
     "x_2")

###
# Fourth: Mean (-5, 10) and no correlation
mu <- c(-5,10)
Sigma <- matrix(c(1, 0, 0, 1), nrow = 2, ncol = 2)
samp <- multinorm(mu = mu, Sigma = Sigma)
plot(samp, asp = 1, main = "Correlation = 0", xlab = "x_1", ylab =
     "x_2")

```



2 Questions to think about

- Can you construct a zero-inflated normal distribution and find a suitable application of it?
- How can you generate samples from an F -distribution?
- Can you generate samples from a Cauchy distribution that has mean 10 (instead of 0)?