Sta-230: Probability **Summer 2020 I**

Key ideas for topic 18: Bivariate Normal Distribution June 15

A pair of random variables (X, Y) are distributed according to the bivariate normal distribution, if any linear combination $aX + bY$ is normally distributed.

A bivariate normal distribution is completely characterized by 5 numbers (parameters) that are $\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho_{XY}$.

Marginal distributions

$$X \sim \text{Normal}(\mu_X, \sigma_X^2)$$
$$Y \sim \text{Normal}(\mu_Y, \sigma_Y^2).$$

The conditional distributions

$$X|Y \sim \text{Normal}(\mu_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(Y - \mu_Y), (1 - \rho_{XY}^2)\sigma_X^2)$$
$$Y|X \sim \text{Normal}(\mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(X - \mu_X), (1 - \rho_{XY}^2)\sigma_Y^2).$$

The normal distribution in the formulae above is that of a univariate normal: if $X \sim \text{Normal}(\mu, \sigma^2)$, then X has the density

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

Multivariate normal distribution (Out of the scope of this course)

More generally, a multivariate normal distribution is a vector-valued distribution of $(X_1, \ldots, X_n)$ and it has parameters mean vector μ of length n and an $n \times n$ covariance matrix $\mathbf{\Sigma}$.

The multivariate normal distribution has the density

$$f_{(X_1, X_2, \ldots, X_n)}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \text{Det}(\Sigma)}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)},$$

where $\mathbf{\Sigma}$ is a symmetric positive definite matrix.

Sta-230: Probability **Summer 2020 I**

Key ideas for topic 19: Inequalities and Law of Large Number June 16

Chebyshev inequality: For a random variable X

$$\mathbf{P}(|x| \geq \varepsilon) \leq \frac{\mathbb{E}x^2}{\varepsilon^2}.$$

Weak law of large numbers: For random variables $X_1, ..., X_n \overset{iid}{\sim} f$ with $\sigma^2 < \infty$ define $S_n = X_1 + X_2 + \cdots + X_n$

$$\begin{aligned} \mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) &\leq \frac{1}{\varepsilon^2}\frac{\sigma^2}{n}, \\ \lim_{n\to\infty} \mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) &= 0, \forall \varepsilon > 0. \end{aligned}$$

Strong law of large numbers: Define the event

$$\mathcal{E} = \left\{\omega \in \Omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\right\},$$

as an infinite random sample $\omega = \{X_1, ..., X_n, ...\}$ for which the sample mean is equal to the true mean. The strong law of large number states that

$$\mathbf{P}(\mathcal{E}) = 1,$$

and holds if $\sigma^2(X_j) < \infty$ and $\mathbb{E}(X_J^4) < \infty$ for all j.

Sta-230: Probability **Summer 2020 I**

Key ideas for topic 20: The Central limit Theorem June 17

Central limit theorem: Given a sequence of $\{X_1, ..., X_n\}$ of iid random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$

$$\lim_{n\to\infty} \frac{\frac{1}{n}\sum_i X_i - \mu}{\sigma/\sqrt{n}} \stackrel{d}{=} \text{N}(0,1).$$

Convergence in probability: A sequence of random variables $\{X_1, ..., X_n\}$ converges to a distribution F if

$$\lim_{n\to\infty} \Pr(X_n < z) = F(z), \ \ \forall z$$

where $F(z)$ is the cdf.

Convergence in probability means the distribution function converges.

Sta-230: Probability **Summer 2020 I**

Key ideas for topic 21: Markov Chains June 18

Markov Chain

A stochastic process X_k with discrete index set $k = 1, \dots$ is a discrete-time Markov Chain if

$$\Pr(X_n | X_1, \dots, X_{n-1}) = \Pr(X_n | X_{n-1})$$

for all n.

Transition vectors

If the random variables X_k take a finite set of values $a_1, \dots, a_n$, the transition distribution $\Pr(X_n | X_{n-1} = a_j)$ is a vector:

$$\mathbf{p_j} := \{p_{1,j}, p_{2,j}, \dots, p_{n,j}\}$$

with $p_{i,j} = \Pr(X_m = a_i | X_{m-1} = a_j)$, and naturally $\sum_{i=1}^{n} p_{i,j} = 1$.

State space The values $a_1, \dots a_n$ are called the states of the process and a Markov process with finitely many states is called a finite-state Markov process.

Transition matrix

Stacking the vectors $\mathbf{p_j}$ together we get the **transition kernel** that we can conveniently express as a matrix:

$$\mathbf{P} = \begin{pmatrix} \mathbf{p_1} \\ \mathbf{p_2} \\ \vdots \\ \mathbf{p_n} \end{pmatrix} = \begin{pmatrix} p_{1,1} & p_{2,1} & \cdots & p_{n,1} \\ p_{1,2} & p_{2,2} & \cdots & p_{n,2} \\ \vdots & & & \\ p_{1,n} & p_{2,n} & \cdots & p_{n,n} \end{pmatrix}$$

The Markov chain is **time homogeneous** if the kernel/matrix doesn't depend on m (No $p_{i,j}$ depends on m).

The transition matrix allows us to compute the marginal distribution $p(X_{m+1})$ of X_{m+1} if we know the distribution $p(X_m)$ of X_m:

$$p(X_{m+1}) = p(X_m) \begin{pmatrix} p_{1,1} & p_{2,1} & \cdots & p_{n,1} \\ p_{1,2} & p_{2,2} & \cdots & p_{n,2} \\ \vdots & & & \\ p_{1,n} & p_{2,n} & \cdots & p_{n,n} \end{pmatrix}$$

which we can compute as vector $p(X_m)$ multiplied (from right) by matrix $\mathbf{P}$.

The kth step transition kernel gives the conditional probability $\Pr(X_{m+k}|X_m)$, and (for a time homogeneous process) it is obtained as a matrix power $\mathbf{P}^k$.

Infinite state space (not in the scope of this course)

For continuous distributions, a Markov Kernel may be expressed as a transition density $K(x, y)$ that satisfies

$$f_{X_m}(x) = \int_{-\infty}^{\infty} K(x, y) f_{X_{m-1}}(y) \; dy.$$

These are used in Gibbs' sampling, which is an example of a *Markov Chain Monte Carlo* method that has very important applications in Bayesian statistics.

Sta-230: Probability **Summer 2020 I**

Key ideas for topic 21: Markov Chains Continued June 19

Stationary Distribution

A discrete-time Markov space with finite state space has a stationary distribution π if

$$\pi = \pi \mathbf{P}$$

for $\mathbf{P}$ the transition matrix.

Aperiodic

A state is B aperiodic, if there exists and N such that $\Pr(X_M = B|X_0 = B) > 0$ holds for all $M \geq N$. The chain is aperiodic if all the states are.

Irreducibility

If for any pair i, j, there exists $m \in \mathbb{Z}_+$ such that $\Pr(X_{k+m} = a_i|X_k = a_j) > 0$. In this case the chain is said to be **irreducible**, which implies the stationary distribution exists and is unique.

If we can find an m such that $\Pr(X_{k+m} = a_i|X_k = a_j) > 0$ hold simultaneously for all pairs (i, j), then the stationary distribution is given by the

$$\pi = \lim_{m \to \infty} p_0 \mathbf{P}^m,$$

for any arbitrary initial state distribution p_0. Such Markov chains are called regular and the transition matrix primitive, and this condition holds whenever the chain is aperiodic and irreducible.