Lecture 2: Probability: many variables

Math 586

January 22, 2009

Recap: (Lecture 1)

- "A *random variable* is a variable whose values are randomly generated according to some probabilistic mechanism" - Isaaks & Srivastava
- X = random variable, x = number.
- $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt$ (cumulative) distribution function.
- $f(x)$ density (continuous) or $P(X = x_i)$ (discrete)

Joint distribution

- If $X_1, ..., X_n$ are r.v. then

$$F(x_1, ..., x_n) = P(X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n)$$

  is their **joint distribution function**.

- If $F(x_1, ..., x_n)$ is differentiable in each x_i then

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n}$$

  is their **joint density function**.

- If $\mathbf{X} = (X_1, X_2, ..., X_n)'$ is a **random vector** (column vector, $n \times 1$), and subset $A \subset \mathbb{R}^n$ then

$$P(\mathbf{X} \in A) = \int\int_A \cdots \int f(x_1, x_2, ..., x_n)\, dx_1\, dx_2 ... dx_n$$

- Expectation of a function

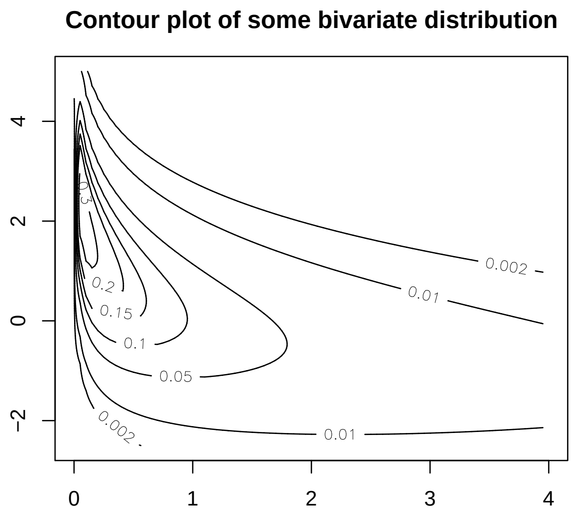
$$\mathbb{E}\left[g(X_1, ..., X_n)\right] = \int\int \cdots \int g(x_1, x_2, ..., x_n) \cdot f(x_1, x_2, ..., x_n)\, dx_1\, dx_2 ... dx_n$$

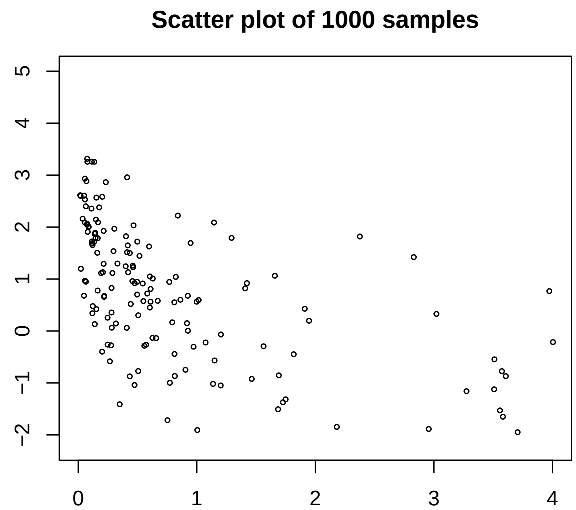
- Statistical independence

  $X_1, ..., X_n$ are **statistically independent** iff

$$f(x_1, x_2, ..., x_n) = f_1(x_1) \cdot f_2(x_2) \cdot .... \cdot f_n(x_n)$$

  One or more r.v.'s can be functionally dependent even though they are statistically independent.





Estimates of Distributions

Conceptual model: Population (of all feasible observations) from which we draw samples to estimate distribution and its properties, such as expected value.

Example: consider a manufactured item with design engineering strength, but actual strength varies in production. The "model strength" is a r.v. X with distribution $F(x)$. We don't know F *a priori* but must estimate it from the data.

- Estimate $F(x_0)$ based on n samples $x_1, x_2, ..., x_n$, e.g. using <u>empirical CDF</u>

$$\hat{F}(x_0) = \frac{\#\{x_i \leq x_0\}}{n} \qquad \text{a.k.a. } \textit{ogive}$$

- Estimate the mean strength by using sample mean

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  This is an estimator for $\mathbb{E}[X]$.

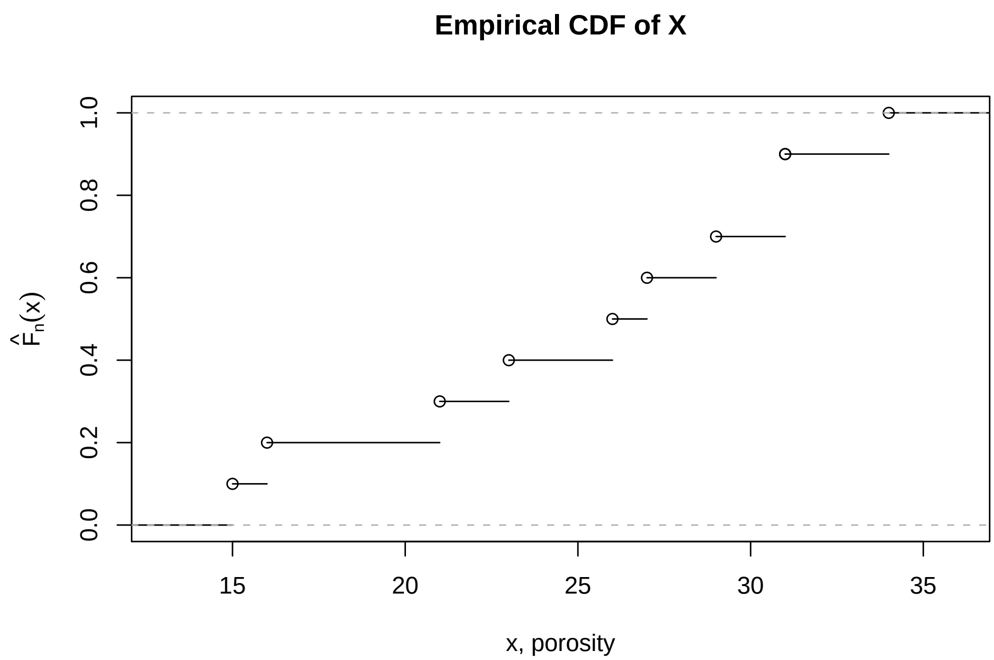
- Note: there is an important difference between the estimate ($\hat{F}$ or $\overline{x}$) and the true value.

- Algorithm to calculate $\hat{F}$: sort x_i's from smallest to largest, get $x_1^*, x_2^*, ..., x_n^*$ then

$$\hat{F}(x) = \frac{j}{n}, \quad x_j^* \leq x \leq x_{j+1}^*$$

example: take $n = 10$ samples of porosity (in %):

$34, 27, 15, 23, 21, 31, 26, 29, 16, 31$ reorder: $\Rightarrow$ $15, 16, 21, 23, 26, 27, 29, 31, 31, 34$



Also, calculate sample mean $\overline{x}$.

Variance and Covariance

- **Variance**

  - Let $\mathbb{E}[X] = \mu$.

$$Var(X) = \sigma^2 = \mathbb{E}[(X - \mu)^2] \quad \text{2nd central moment}$$

  - σ is the Standard Deviation
  - Also $Var(X) = \mathbb{E}[X^2 - 2\mu X + \mu^2] = \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \Rightarrow$

$$Var(X) = \mathbb{E}[X^2] - \mu^2 \qquad \text{"Computational formula for variance"}$$

- **Covariance**
  - Given r.v.'s X_1 and X_2 with means μ_1, μ_2,
  $$Cov(X_1, X_2) = \mathbb{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] = \mathbb{E}\left[X_1 X_2\right] - \mu_1 \mu_2$$
  $$\texttt{Note:} \quad Cov(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y)\, dx\, dy$$
  Replace integral by summation for discrete case.
  - If X_1 and X_2 are statistically independent then
  $$Cov(X_1, X_2) = 0$$
  - **Correlation coefficient** between X_1 and X_2 with st.dev. σ_1, σ_2
  $$\rho = \frac{Cov(X_1, X_2)}{\sigma_1 \sigma_2}$$
- **Variance of the sum**:
$$Var(a_1 X_1 + a_2 X_2) = \mathbb{E}\left[(a_1 X_1 - a_1 \mu_1 \ + \ a_2 X_2 - a_2 \mu_2)^2\right] =$$
$$= a_1^2 \mathbb{E}\left[(X_1 - \mu_1)^2\right] + a_2^2 \mathbb{E}\left[(X_2 - \mu_2)^2\right] + 2 a_1 a_2 \mathbb{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] =$$
$$= a_1^2\, Var(X_1) + a_2^2\, Var(X_2) + 2 a_1 a_2 Cov(X_1, X_2)$$
Thus,
$$Var(a_1 X_1 + a_2 X_2) = a_1^2\, Var(X_1) + a_2^2\, Var(X_2) + 2 a_1 a_2 Cov(X_1, X_2)$$
  - If X_1, X_2 are independent, then $Cov = 0$ and
  $$Var(a_1 X_1 + a_2 X_2) = a_1^2 Var(X_1) + a_2^2 Var(X_2)$$
  - Consider n r.v.'s, $X_1, ..., X_n$ then
  $$\boxed{Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2\, Var(X_i) + \sum_{i} \sum_{j \neq i} a_i a_j\, Cov(X_i, X_j)}$$

  Note: kriging algorithms are based on minimizing variance of linear combinations of r.v.'s. This expression for variance is very important. The covariance on RHS will carry information about spatial continuity.
  - If $X_1, X_2, ..., X_n$ are independent, then $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2\, Var(X_i)$