Random Samples

November 5, 2009

1 Basic Concepts

The random variables $X_1, X_2, \dots, X_n$ are called a **a random sample** or **independent and identically distributed** if they are independent and have a common distribution.

Let f denote the density (mass) function of a single random variable in the sample Consequently, the joint density (mass) function is the product of the marginal density (mass) functions

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f(x_1)\cdots f(x_n) = \prod_{i=1}^n f(x_i).$$

A set of observations $x_1, x_2, \dots, x_n$ from this distribution is called the **data**.

If we are examining a member of a parametric family of random variables, then we migth write

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n|\theta) = f(x_1|\theta)\cdots f(x_n|\theta) = \prod_{i=1}^n f(x_i|\theta).$$

Example 1. • *For Bernoulli random variables with parameter p,*

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n|\theta) = \prod_{i=1}^n p^{x_1}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_1}(1-p)^{1-\sum_{i=1}^n x_i}.$$

- *For normal random variables with mean μ and variance σ^2,*

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x_i-\mu)^2}{2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp -\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2.$$

A summary $T(x_1, x_2, \dots, x_n)$ of the data is called a **statistic**. The distribution of the random variable $T(X_1, X_2, \dots, X_n)$ is called the **sampling distribution**. *Note that the statistic cannot be a function of the parameter.*

Example 2 (statistics)**.** • *The maximum*

$$T(x_1, x_2, \dots, x_n) = \max\{x_1, x_2, \dots, x_n\}.$$

- *The* **sample mean** *is the arithmetic average of the data.*

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

- *A* **weighted sample mean** *with weights* w_i *is*

$$T(x_1, x_2, \ldots, x_n) = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n}.$$

- *The* **sample variance**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

2 Sums of Random Variables

We begin with a little algebra.

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - a)^2 &= \sum_{i=1}^{n}((x_i - \bar{x}) + (\bar{x} - a))^2 \sum_{i=1}^{n}(x_i - \bar{x})^2 + 2\sum_{i=1}^{n}(x_i - \bar{x})(\bar{x} - a) + \sum_{i=1}^{n}(\bar{x} - a))^2 \\
&= \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(\bar{x} - a)^2
\end{aligned}$$

because the cross term equals to zero.

Take $a = \bar{x}$ to see that

$$\min_a \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Take $a = 0$ to see that

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right).$$

For the random sample, $X_1, X_2, \ldots, X_n$, we also have the identities

1. $E\left[\sum_{i=1}^{n} g(X_i)\right] = nEg(X_1)$ provided that $Eg(X_1)$ exists.
2. $\mathrm{Var}\left(\sum_{i=1}^{n} g(X_i)\right) = n\mathrm{Var}(g(X_1))$ provided that $\mathrm{Var}(g(X_1))$ exists.

For a random sample with mean μ and variance σ^2, these identities lead us to

1. $E\bar{X} = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}n\mu = \mu$.
2. $\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$.

3. The mean of the sample variance

$$\begin{aligned} ES^2 &= \frac{1}{n-1}E\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right] = \frac{1}{n-1}(nEX_1^2 - nE\bar{X}^2) \\ &= \frac{n}{n-1}\left((\sigma^2+\mu^2) - \left(\frac{\sigma^2}{n}+\mu^2\right)\right) = \frac{n}{n-1}\left(1-\frac{1}{n}\right)\sigma^2 = \sigma^2. \end{aligned}$$

If we have that

$$E_\theta[T(X_1, X_2, \ldots, X_n)] = k(\theta),$$

then we say that $T(x_1, x_2, \ldots, x_n)$ is an **unbiased estimator** of $k(\theta)$. Thus,

1. $\bar{x}$ is an unbiased estimator of the mean μ.
2. s^2 is an unbiased estimator of the variance σ^2.

For the moment generating function of $\bar{X}$,

$$\begin{aligned} M_{\bar{X}}(t) &= E[\exp t\bar{X}] = E[\exp \frac{t}{n}(X_1 + X_2 + \cdots + X_n)] \\ &= E[\exp \frac{t}{n}X_1]E[\exp \frac{t}{n}X_2]\cdots E[\exp \frac{t}{n}X_n] = M_{X_1}\left(\frac{t}{n}\right)^n \end{aligned}$$

Example 3. *1. For a sample of normal random variables with mean μ and variance σ^2,*

$$M_{X_1}(t) = \exp t\mu + \frac{t^2}{2}\sigma^2.$$

Thus,

$$M_{\bar{X}} = \left(\exp \frac{t}{n}\mu + \frac{t^2}{2n^2}\sigma^2\right)^n = \exp t\mu + \frac{t^2}{2n}\sigma^2$$

and $\bar{X}$ is normal, mean μ and variance σ^2/n.

2. For a sample of gamma random variables with parameters α and β

$$M_{X_1}(t) = (1-\beta t)^{-\alpha}.$$

Thus,

$$M_{\bar{X}} = (1-\beta\frac{t}{n})^{-n\alpha}$$

and $\bar{X}$ is a gamma random variable with parameters $n\alpha$ and β/n