What is a probability density function ?

Definition

A random variable is a variable whose value is determined by the outcome of a random experiment. A random variable is said discrete if its values are countable, or continuous otherwise.

Definition

The probability density function (p.d.f.) f of a random variable X gives a natural description of the distribution of X and allows probabilities associated with X to be computed:

$$P(a < X < b) = \int_a^b f(x)\ dx \quad \forall (a,b)\ a < b$$

Parameters and Moments of a probability density function I

Definition

A parameter, θ, is a function of the probability density function (p.d.f.) f e.g.:

$$\theta = t(f)$$

Definition

A moment of order n is a parameter of the probability density function (p.d.f.) f, defined as:

$$\theta = \int x^n\ f(x)\ dx \quad \text{raw moment}$$

Central moments are also defined using the first raw moment $\mu = \int x\ f(x)\ dx$:

$$\theta = \int (x-\mu)^n\ f(x)\ dx \quad \text{central moments } (n > 1)$$

Parameters and Moments of a probability density function II

if θ is the mean

$$\theta = \mathbb{E}_f(x) = \int_{-\infty}^{+\infty} x\ f(x)\ dx = \mu_f$$

if θ is the variance

$$\theta = \mathbb{E}_f[(x-\mu_f)^2] = \int_{-\infty}^{+\infty} (x-\mu_f)^2\ f(x)\ dx = \sigma_f^2$$

Parameters and Moments of a probability density function III

Exercise. Compute the mean and variance of the following distributions:

1. Dirac distribution
$$f(x) = \delta(x-6)$$
2. The normal distribution [1]
$$f(x) = \frac{1}{2\sqrt{2\pi}} \exp -\frac{(x-1)^2}{8}$$
3. $$f(x) = 0.2\ \delta(x-1) + 0.8\ \delta(x-6)$$
4. $$f(x) = 0.2\ \mathcal{N}(\mu{=}1,\sigma{=}2) + 0.8\ \mathcal{N}(\mu{=}6,\sigma{=}1)$$

[1] also noted $\mathcal{N}(\mu = 1, \sigma = 2)$

Estimation $\hat{f}$ of a density function f

Lets assume that we have a set of samples or observations $\{x_i\}_{i=1,\cdots,n}$ of the random variable X.

We can differentiate two approaches to estimate the p.d.f. $f(x)$:

- parametric
- non-parametric

Today we focus on non-parametric approaches

Non-Parametric estimation: Empirical p.d.f.

Definition

The empirical density function $\hat{f}(.)$ is computed using a set of samples $\{x_i\}_{i=1,\cdots,n}$ such that:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^n \delta(x-x_i)$$

where $\delta(\cdot)$ is the Dirac delta function.

Non-Parametric estimation: Histograms I

Definition (Histogram)

Lets consider a set of observations $\{x_i\}_{i=1,\cdots,n}$ of the random variable X. A histogram defined as:

$$\hat{f}(x) = \frac{1}{nh}(\text{no. of } x_i \text{ in the same bin as } x)$$

is an estimate of the probability density function f. Note that we need to specify the origin x_0 and a bin width h to define the bins of the histogram to be $[x_0 + mh, x_0 + (m+1)h]$ with $m \in \mathbb{Z}$.

Exercise: Propose a procedure to compute the histogram of a grey-level image.

Non-Parametric estimation: Kernel density I

Definition

The kernel estimator of a probability density function is defined as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right)$$

h is called the bandwidth, and $k(\cdot)$ is the kernel function which satisfies:

$$\int_{-\infty}^{+\infty} k(x)\; dx = 1$$

Example (kernels)

$$k(x) = \begin{cases} 1/2 & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

Exercise

Consider the observations $\{94; 197; 16; 38; 99; 141; 23\}$ of a r.v. X.

1. Draw their empirical p.d.f , their histogram, their kernel distribution with a gaussian kernel.
2. Compare and comment those different estimates of the p.d.f.

Estimates of parameters I

Definition

An estimate $\hat{\theta}$ of the parameter $\theta = t(f)$ is a function of the estimated p.d.f. $\hat{f}$ or the sample $\mathbf{x} = \{x_i\}$, e.g.:

$$\hat{\theta} = t(\hat{f})$$

or also written $\hat{\theta} = s(\mathbf{x})$.
The Plug-in estimate $\hat{\theta} = t(\hat{f})$ is computed using the empirical p.d.f..

Estimates of parameters II

Plug-in estimate of the mean

$$\begin{aligned} \hat{\theta} \;= t(\hat{f}) &= \int_{-\infty}^{+\infty} x\; \hat{f}(x) dx \\ &= \int_{-\infty}^{+\infty} x\; \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \delta(x - x_i)\; dx \\ &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} x_i \\ &= s(\mathbf{x}) = \overline{x} \end{aligned}$$

Exercise: compute the plug-in estimate of the variance.

Example: Difference between θ and $\hat{\theta}$ I

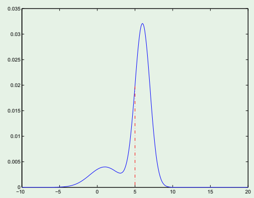
Computing the mean knowing f

Lets assume we know the p.d.f. f:

$$f(x) = 0.2\; \mathcal{N}(\mu{=}1, \sigma{=}2) + 0.8\; \mathcal{N}(\mu{=}6, \sigma{=}1)$$

Then the mean is computed:

$$\begin{aligned} \mu_f = \mathbb{E}_f(x) &= \int_{-\infty}^{+\infty} x\; f(x)\; dx \\ &= 0.2 \cdot 1 + 0.8 \cdot 6 \\ &= 5 \end{aligned}$$



Example: Difference between θ and $\hat{\theta}$ II

Estimating the mean knowing the observations $\mathbf{x}$

Observations $\mathbf{x} = (x_1, \cdots, x_{100})$:

7.0411	4.8397	5.3156	6.7719	7.0616
5.2546	7.3937	4.3376	4.4010	5.1724
7.4199	5.3677	6.7028	6.2003	7.5707
4.1230	3.8914	5.2323	5.5942	7.1479
3.6790	0.3509	1.4197	1.7585	2.4476
−3.8635	2.5731	−0.7367	0.5627	1.6379
−0.1864	2.7004	2.1487	2.3513	1.4833
−1.0138	4.9794	0.1518	2.8683	1.6269
6.9523	5.3073	4.7191	5.4374	4.6108
6.5975	6.3495	7.2762	5.9453	4.6993
6.1559	5.8950	5.7591	5.2173	4.9980
4.5010	4.7860	5.4382	4.8893	7.2940
5.5741	5.5139	5.8869	7.2756	5.8449
6.6439	4.5224	5.5028	4.5672	5.8718
6.0919	7.1912	6.4181	7.2248	8.4153
7.3199	5.1305	6.8719	5.2686	5.8055
5.3602	6.4120	6.0721	5.2740	7.2329
7.0912	7.0766	5.9750	6.6091	7.2135
4.9585	5.9042	5.9273	6.5762	5.3702
4.7654	6.4668	6.1983	4.3450	5.3261

From the samples, the mean can be computed:

$$\bar{x} = \frac{\sum_{i=1}^{100} x_i}{100} = 4.9970$$

Accuracy of arbituary estimates $\hat{\theta}$ I

We can compute an estimate $\hat{\theta}$ of a parameter θ from an observation sample $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. But

how accurate is $\hat{\theta}$ compared to the real value θ ?

Our attention is focused on questions concerning the probability distribution of $\hat{\theta}$. For instance we would like to know about

- its standard error
- its confidence interval
- etc.

In this course, only the concept of standard error is introduced.

Accuracy of arbituary estimates $\hat{\theta}$ II

Definition

The standard error is the standard deviation of a statistic $\hat{\theta}$. As such, it measures the precision of an estimate of the statistic of a population distribution.

$$se(\hat{\theta}) = \sqrt{var_f[\hat{\theta}]}$$

Standard error of $\bar{x}$

We have:

$$\mathbb{E}_f\left[(\bar{x} - \mu_f)^2\right] = \frac{\sum_{i=1}^{n} \mathbb{E}_f\left[(x_i - \mu_f)^2\right]}{n^2} = \frac{\sigma_f^2}{n}$$

Then

$$se_f(\bar{x}) = [\mathrm{var}_f(\bar{x})]^{1/2} = \frac{\sigma_f}{\sqrt{n}}$$

Accuracy of arbituary estimates $\hat{\theta}$ III

Suppose now that f is unknown and that only the random sample $\mathbf{x} = (x_1, \cdots, x_n)$ is known. As μ_f and σ_f are unknown, we can use the previous formula to compute a plug-in estimate of the standard error.

Definition

The estimated standard error of the estimator $\hat{\theta}$ is defined as:

$$\hat{\mathrm{se}}(\hat{\theta}) = \mathrm{se}_{\hat{f}}(\hat{\theta}) = [\mathrm{var}_{\hat{f}}(\hat{\theta})]^{1/2}$$

Estimated standard error of $\bar{x}$

$$\hat{\mathrm{se}}(\bar{x}) = \frac{\hat{\sigma}}{\sqrt{n}}$$

Example on the mouse data

Data (Treatment group)	94; 197; 16; 38; 99; 141; 23
Data (Control group)	52; 104; 146; 10; 51; 30; 40; 27; 46

Table: The mouse data [Efron]. 16 mice assigned to a treatment group (7) or a control group (9). Survival in days following a test surgery.

Did the treatment prolong survival ?

Example on the mouse data

Mean and Standard error for both groups

	$\bar{x}$	$\hat{\mathrm{se}}$
Treatment	86.86	25.24
Control	56.22	14.14

Conclusion at first glance

It seems that mice having the treatment survive $d = 86.86 - 56.22 = 30.63$ days more than the mice from the control group.

Example on the mouse data

Stantard error of the difference $d = \overline{x}_{Treat} - \overline{x}_{Cont}$

$\overline{x}_{Treat}$ and $\overline{x}_{Cont}$ are independent, so the standard error of their difference is $\hat{\text{se}}(d) = \sqrt{\hat{\text{se}}^2_{Treat} + \hat{\text{se}}^2_{Cont}} = 28.93$. We see that:

$$\frac{d}{\hat{\text{se}}(d)} = \frac{30.63}{28.93} = 1.05$$

This shows that this is an insignificant result as it could easily have arised by chance (i.e. if the test was reproduced, it is *likely possible* to measure datasets giving $d = 0$!).
Therefore, we can not conclude with certainty that the treatment improves the survival of the mice.