09.03.2016

Joint distributions/marginal distributions

Independencs of RVs

multdim RV:

$$(X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$$

add pic

$$[(X_1, \dots, X_n) \in B] = (X_1, \dots, X_n)^{-1}(B) \in \mathcal{A}$$

$$P_{X_1,\dots,X_n}(B) = \mathbb{P}[(X_1, \dots, X_n) \in B] \tag{1}$$

is a probability measure on $\mathbb{R}^n$: joint distribution of $X_1, \dots, X_n$.

P_{X_k} probability on $\mathbb{R}$. It is called the k-th marginal of $P_{X_1,\dots,X_k}$

add pic

$P_{X_k}(I) = \mathbb{P}[X_k \in I] = \mathbb{P}[(X_1, \dots, X_k) \in \mathbb{R} \times \mathbb{R} \times \dots \times \underbrace{I}_{k} \times R \times \mathbb{R}]$ where $I \subset \mathbb{R}$ interval.

discrete:

$(X_1, \dots, X_n)$ takes at most countable many values in $\mathbb{R}^d$.

joint discrete density: $p_{X_1,\dots,X_n}(x_1, \dots, x_n) = \mathbb{P}[X_1 = x_1, \dots, X_n = x_n]$

$$\sum_{(x_1,\dots,x_n)} p_n(x_1, \dots, x_n) = t$$

$$\mathbb{P}[(X_1, \dots, X_n) \in B] = \sum_{(x_1,\dots,x_n) \in B} p_n(x_1, \dots, x_n)$$

$$p_{X_k}(a) = \sum_{(x_1,\dots,x_n) \in \mathbb{R}^n, x_k = a} p_n(x_1, \dots, x_n)$$

often two RVs: (X, Y). Then

$$p_{X,Y}(x, y) = \mathbb{P}[X = x, Y = y]$$

$$p_Y(y) = \sum_{x \in \mathbb{R}} p_{X,Y}(x, y)$$

continuous: similar

$$\mathbf{??} = \int \dots \int_B \underbrace{f_{X_1,\dots,X_n}(x_1, \dots, x_n)}_{\text{joint density}} dx_1 \dots dx_n$$

k-th marginal:

$$f_{X_k}(x_k) =$$

add pic

Definition. *RVs $X_1, \dots, X_n$ are called independent if for all intervals (Borel sets in $\mathbb{R}$) $I_1, \dots, I_n \subset \mathbb{R}$*

$$[X_1 \in I_1], \dots, [X_n \in I_n]$$

are independent. Then

$$\mathbb{P}[X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n] = \mathbb{P}[X_1 \in I_1] \dots \mathbb{P}[X_n \in I_n]$$

for discrete/continuous cases:

$$p_{X_1,\dots,X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n)$$

The same holds for f almost everywhere.

1 Variance: Markov- and Chebyshev inequalities

Recall:

$$\mathbb{E}(X) = \int_\Omega X d\,\mathbb{P} = \begin{cases} \sum_x xp(x) \\ \int_\mathbb{R} xf(x)dx \end{cases}$$

Definition. *Let X be a random variable such that $\mathbb{E}(X)$ exists and is finite. Then the variance is defined as*

$$\mathbb{V}(X) \coloneqq \mathbb{E}((X - \mathbb{E}(X))^2)$$

add pic

Recall: If X_1, X_2 are RVs with finite expectation, then

$$\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 \mathbb{E}(X_1) + \lambda_2 \mathbb{E}(X_2)$$

If $X_1 \geq 0$ a.s., then $\mathbb{E}(X_1) \geq 0$. Now X RV:

$$g : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$$

measurable (good, e.g. continuous)

$$g(X)(\omega) = g \circ X(\omega) = g(X(\omega))$$

Example. $\mathbb{E}((X - \mathbb{E}(X))^2)$, *then* $g(x) = (x - \mu)^2$.

If $\mathbb{E}(g(X))$ is finite, then

$$= \int_\mathbb{R} g(x) dP_X(x)$$

$$= \begin{cases} \sum_{x \in \mathbb{R}} g(x) p_X(x) & \text{discrete} \\ \int_\mathbb{R} g(x) f_X(x) dx & \text{continuous} \end{cases}$$

Now consider:

$$\begin{aligned}
(X - \mathbb{E}(X))^2 &= X^2 - 2\,\mathbb{E}(X)X + \mathbb{E}(X)^2 \\
&= X^2 - 2\mu X + \mu^2 \\
\mathbb{V}(X) &= \mathbb{E}(X^2) - 2\mu\,\mathbb{E}(X) + \mu^2 \\
&= \mathbb{E}(X^2) - 2\mu\mu + \mu^2 \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2
\end{aligned}$$

Let (X, Y) be a 2-dim RV with $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ finite. Then we know

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

. Consider now the variance:

$$\begin{aligned}
\mathbb{V}(X + Y) &= \mathbb{E}((X + Y - \mathbb{E}(X) - \mathbb{E}(Y))^2) \\
&= \mathbb{E}((X - \mathbb{E}(X))^2) + \mathbb{E}((Y - \mathbb{E}(Y))^2) + 2\,\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\
&= \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathrm{Cov}(X, Y)
\end{aligned}$$

Important fact: if X, Y are independent (and the involved expected values are finite), then

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\,\mathbb{E}(Y)$$

discrete case:
$(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$

$$\begin{aligned} g &: \mathbb{R}^2 \to \mathbb{R} \\ g(x, y) &= x \cdot y \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}(g(X,Y)) &= \sum_{(x,y)\in\mathbb{R}^2} g(x,y) \cdot p_{X,Y}(x,y) \\ &= \sum_x \sum_y xy p_X(x) p_Y(y) \\ &= \sum_x x p_X(x) \sum_y y p_Y(y) \\ &= \mathbb{E}(X)\,\mathbb{E}(Y) \end{aligned}$$

Formula in independent case:

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$$
$$\mathbb{E}(X - \mathbb{E}(X))\,\mathbb{E}(Y - \mathbb{E}(Y)) = 0$$

Trap for exam:

$$\mathbb{V}(X - Y) = \mathbb{V}(X) + \mathbb{V}(-Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

Lemma 1.1 (Markov inequality)**.** *Let X be a non-negative real random variable. And $0 < \mathbb{E}(X) < \infty$. Then*

$$\mathbb{P}[X \geq a\,\mathbb{E}(X)] \leq \frac{1}{a}$$

for all $a > 0$.

Proof. Let $I = \not{\Vdash}_{[X \geq a\,\mathbb{E}(X)]}$ be a random variable. $I(\omega) = \begin{cases} 1 & X(\omega) \geq a\,\mathbb{E}(X) \\ 0 & \text{otherwise} \end{cases}$

$$\mathbb{E}(I) = \mathbb{P}[X \geq a\,\mathbb{E}(X)]$$

$$X \geq X \cdot I \geq a\,\mathbb{E}(X) \cdot I$$

$$\mathbb{E}(X) \geq \mathbb{E}(a\,\mathbb{E}(X) I) = a\,\mathbb{E}(X) \cdot \mathbb{E}(I) = a\,\mathbb{E}(X)\,\mathbb{P}[X \geq a\,\mathbb{E}(X)]$$

Then

$$\frac{1}{a} \geq \mathbb{P}[X \geq a\,\mathbb{E}(X)]$$

$$j$$

$\square$

Remark (Exercise)**.** *Let* $X \geq 0$ *a.s. and* $\mathbb{E}(X) = 0$*. Then* $X = 0$ *a.s.*

Corollary 1.2 (Chebyshev inequality)**.** *Let* X *be an arbitrary random variable such that* $\mathbb{E}(X)$*,* $\mathbb{V}(X)$ *are finite. Then for any* $a > 0$

$$\mathbb{P}[(X - \mathbb{E}(X)) \geq a] \leq \frac{\mathbb{V}(X)}{a^2}$$

Proof. Let $Y \coloneqq (X - \mathbb{E}(X))^2 \geq 0$.

$$\begin{aligned}
\mathbb{P}[(X - \mathbb{E}(X)) \geq a] &= \mathbb{P}[(X - \mathbb{E}(X))^2 \geq a] \\
&= \mathbb{P}[Y \geq \frac{a^2 \, \mathbb{E}(Y)}{\mathbb{E}(Y)}] \\
&= \mathbb{P}[Y \geq \frac{a^2}{\mathbb{E}(Y)} \, \mathbb{E}(Y)] \\
&= \mathbb{P}[Y \geq \frac{a^2}{\tilde{a}} \, \mathbb{E}(Y)] \\
&\geq \frac{\mathbb{E}(Y)}{a^2} \\
&= \frac{\mathbb{V}(X)}{a^2}
\end{aligned}$$

□

Theorem 1.3 (weak law of large numbers)**.** *Let* $(X_n)_{n \in \mathbb{N}}$ *be a sequence of iid (independent identically distributed) random variables with* $\mu = \mathbb{E}(X_n)$ *and* $\sigma^2 = \mathbb{V}(X_n)$ *finite and* $\sigma^2 > 0$*.*

$$\overline{X_n} = \frac{1}{n}(X_1 + \cdots + X_n)$$

Then

$$\overline{X_n} \to \mu$$

in probability.

Proof. $\mathbb{E}(\overline{X_n}) = \mu$

$$\begin{aligned}
\mathbb{V}(\overline{X_n}) &= \frac{1}{n^2} \, \mathbb{V}(X_1 + \cdots + X_n) \\
&= \frac{1}{n^2}(\mathbb{V}(X_1) + \cdots + \mathbb{V}(X_n)) \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}[\left|\overline{X_n} - \mu\right| \geq a] &\leq \frac{\mathbb{V}(\overline{X_n})}{a^2} \\
&= \frac{\sigma^2}{na^2} \to 0
\end{aligned}$$

for all a. □