Probabilistic Modeling and Statistical Computing Fall 2015

September 24, 2015

Expected value of a random variable

Informal definition

$$\mathcal{E}(X) = \sum_{x \in \mathcal{R}} x \cdot p(x)$$

Average outcome of observing X many times

Some properties

$$\mathcal{E}(\alpha X) = \alpha \mathcal{E}(X), \quad \mathcal{E}(X_1 + X_2) = \mathcal{E}(X_1) + \mathcal{E}(X_2)$$

Formal definition, discrete case

$$\mathcal{E}(X) = \sum_{x \in \mathcal{R}} x \cdot p(X = x)$$

Example: Binomial $B(n, p)$ distribution

$$\mathcal{E}(X) = \sum_{i=0}^{n} i \cdot p(X = i) = \sum_{i=0}^{n} i \cdot \binom{n}{i} p^i (1-p)^{n-i} = np\,.$$

Example: Geometric distribution with p

$$\mathcal{E}(X) = \sum_{i=0}^{\infty} i \cdot p(X = i) = \sum_{i=0}^{\infty} i \cdot p(1-p)^i = \frac{1}{p}\,.$$

Formal definition, continuous case

$$\mathcal{E}(X) = \int_{\mathcal{R}} x \cdot p(x) dx$$

Example: Uniform distribution on (a, b)

$$\mathcal{E}(X) = \int_a^b x \cdot \frac{1}{b-a} dx = \cdots = \frac{a+b}{2}.$$

Example: Exponential distribution with **intensity** λ - **note change in notation**

$$\mathcal{E}(X) = \int_0^\infty x \cdot \lambda e^{-x \cdot \lambda} dx = \cdots = \frac{1}{\lambda}$$

Verify this with simulations!

Moments

We can try to compute $\mathcal{E}(f(X))$ where f is a general function.
If $f(x) = x^k$, the result is called a **moment**.

Example: Second moment of a binomial distribution

$$\begin{aligned}
\mathcal{E}(X^2) &= \sum_{i=0}^{n} i^2 \cdot p(X = i) \\
&= \sum_{i=0}^{n} i^2 \cdot \binom{n}{i} p^i (1-p)^{n-i} \\
&= np \cdot (1 - p + np).
\end{aligned}$$

Variance and standard deviation

Don't worry, for most standard probability distributions moments are well known and tabulated.

Suppose $\mu = \mathcal{E}(X)$. Then

$$var(X) = \mathcal{E}((X - \mu)^2)$$

is called the variance. Computational shortcut:

$$var(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2$$

Standard deviation:

$$s(X) = \sqrt{var(X)}$$

Examples

- Binomial distribution $B(m, p)$

$$\mathcal{E}(X) = np$$
$$var(X) = np(1-p)$$
$$s(X) = \sqrt{np(1-p)}$$

- Exponential distribution with intensity λ:

$$\mathcal{E}(X) = \frac{1}{\lambda},\ var(X) = \frac{1}{\lambda^2},\ s(X) = \frac{1}{\lambda}$$

More examples

Suppose X_1 and X_2 both have exponential distributions with parameters $\lambda = 2$ and are independent.

- What is $\mathcal{E}(X_1)$? What is $\mathcal{E}(X_1 + X_2)$? What is $var(X_1 + X_2)$?
- Set up the integral for $\mathcal{E}(\frac{1}{X_1})$. Do you think that's finite? Check with a simulation.
- What would be needed to set up an integral for $\mathcal{E}(\frac{1}{X_1+X_2})$? Use a simulation instead.
- Do you think $\frac{1}{X_1+X_2}$ has a finite standard deviation? How can we check?

Conditional probability

Suppose A and B are events, $prob(B) > 0$.

Definition I

The conditional probability $\mathcal{P}(A|B)$ is defined as

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}$$

The probability given to A if B has occurred.

Example: Roll a die once. Let X be the result.

$$\mathcal{P}(X > 3|X > 2) = \frac{2}{3}, \quad \mathcal{P}(X > 3|X < 5) = \frac{1}{4}.$$

Independence

Suppose A and B are events. The events are called **independent** if

$$\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)\,.$$

If $\mathcal{P}(B) > 0$, this means

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B}{\mathcal{P}(B)} = \mathcal{P}(A)$$

and if $\mathcal{P}(A) > 0$, it also means $\mathcal{P}(B|A) = \mathcal{P}(B)$.

Knowledge about one event does not contain information about the other event.

Example: Roll a die once. A is $X > 2$, B is "X is

Example

Consider n independent trials, success probability p. Let $S_n =$number of successes and let $S_m =$ number of successes in the first m trials **of the same experiment**.

What is $\mathcal{P}(S_m = j | S_n = k)$?

$$\mathcal{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Note: $S_n = k, S_m = j$ means that j successes in m trials are followed by $k - j$ successes in the $n - m$ trials. So ...

$$\mathcal{P}(S_m = j, S_n = k) =$$

$$\binom{m}{j} p^j (1-p)^{m-j} \cdot \binom{n-m}{k-j} p^{k-j} (1-p)^{n-m-(k-j)}$$

and therefore

$$\mathcal{P}(S_m = j, S_n = k) = \frac{\mathcal{P}(S_m = j, S_n = k)}{\mathcal{P}(S_n = k)}$$

$$= \frac{\binom{n}{k}}{\binom{m}{j}\binom{n-m}{k-j}}$$

Hypergeometric distribution, independent of success probability p.

Total probability

Suppose A is some event and $B_1, \ldots, B_n$ are mutually exclusive events that make up the whole sample space $\mathcal{S}$. Then

$$\mathcal{P}(A) = \mathcal{P}(A|B_1) \cdot \mathcal{P}(B_1) + \ldots + \mathcal{P}(A|B_n) \cdot \mathcal{P}(B_n)$$

Example: Assume
$\mathcal{P}(\text{lawn is wet} \mid \text{it has rained}) = .9$
$\mathcal{P}(\text{lawn is wet} \mid \text{it hasn't rained}) = .2$
$\mathcal{P}(rain) = .3$. What is $\mathcal{P}(\text{lawn is wet})$?

Reversing the conditioning

- Often, B is a "cause" and A is an "effect".
- We know $\mathcal{P}(A|B)$ from a "forward model" (cause leading to effect).
- We observe the effect B and would like to know whether A was responsible .
- That is, we want to compute $\mathcal{P}(B|A)$.

$$\boxed{\mathcal{P}(B|A) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(A)} = \mathcal{P}(A|B)\frac{\mathcal{P}(B)}{\mathcal{P}(A)}}$$

Example:

$$\mathcal{P}(\text{it has rained} \mid \text{lawn is wet}) = .9 \cdot \frac{.3}{.41} \approx .66$$

Example: Prosecutor's fallacy

In a city of a million people, somebody commits a crime.
10 people match the description of the criminal.
One of the 10 is charged with the crime.
Let $M =$ "this person matches the description", $I =$ "this person is innocent". Then $\mathcal{P}(M|I) = \frac{9}{999{,}999} \approx 10^{-5}$.
The prosecutor says "if this person were innocent, then a match would be unlikely, thus he is guilty".
This is a fallacy. We must compute $\mathcal{P}(I|M)$, not $\mathcal{P}(M|I)$, and $P(I|M) = \frac{9}{10}$.

Example: Learning from data

A box contains N numbered balls, $N \in \{10, 20, 30\}$.
We don't know N and hence assume that $\mathcal{P}(N = k) = \frac{1}{3}$. Let A_k be the event $N = k$.
Draw two balls with replacement. We observe 14 and 17. Let B be this event.
Learn about N: Replace $\mathcal{P}(A_{20}) = \frac{1}{3}$ with updated $\mathcal{P}(A_{20}|B)$.

$$\mathcal{P}(A_{20}|B) = \mathcal{P}(B|A_{20}) \cdot \frac{\mathcal{P}(A_{20})}{\mathcal{P}(B)}$$

Example: Learning from data 2

$$
\begin{aligned}
\boxed{\mathcal{P}(A_{20}|B)} &= \frac{\mathcal{P}(B|A_{20}) \cdot \mathcal{P}(A_{20})}{\mathcal{P}(B)} \\
&= \frac{\mathcal{P}(B|A_{20}) \cdot \mathcal{P}(A_{20})}{\mathcal{P}(B|A_{20}) \cdot \mathcal{P}(A_{20}) + \mathcal{P}(B|A_{30}) \cdot \mathcal{P}(A_{30})} \\
&= \frac{\frac{1}{400} \cdot \frac{1}{3}}{\frac{1}{400} \cdot \frac{1}{3} + \frac{1}{900} \cdot \frac{1}{3}} = \boxed{\frac{9}{13}}
\end{aligned}
$$

Conditional distribution

Suppose X is a r.v. and B is an event, $\mathcal{P}(B) > 0$.

Definition II

The conditional cdf $F_{X|B}$ is defined as

$$F_{X|B}(x) = \mathcal{P}(X \leq x|B) = \frac{\mathcal{P}(\{X \leq x\} \cap B)}{\mathcal{P}(B)}$$

The probability that $X \leq x$ if B has occurred.

Example: Suppose $X \sim exp(\lambda)$. Find the empirical cdf $F_{X|X \geq x_0}$ with a simulation!

What is $Y = X|B$?

If we can compute $F_{X|B}$, what is the random variable for which this is the cdf?

Simulation approach suggests an answer:

To observe $Y = X|B$, observe X and check if B occurs. **If yes, set** $Y = X$, otherwise repeat.

Example: Simulate $Y = X|1 < X < 2$ where $X \sim N(0, 1)$. Find the expected value.

How many trials are needed to get one observation of Y?

Expected value and conditioning

We can define $Y|B$ and therefore also $\mathcal{E}(Y|B)$. In simulations, this can be computed with subsetting.

Example: Toss a ball 5 times at a target, success probability $p = .4$. If you get 0 successes, toss 5 more times. **What is the expected number of successes?**

$X \sim B(5, .4)$ is number of successes in five tosses, Y = total number of successes.

$$\mathcal{E}(Y) = \mathcal{E}(Y|X = 0) \cdot \mathcal{P}(X = 0) + \mathcal{E}(Y|X > 0) \cdot \mathcal{P}(X > 0)$$

A Bayesian network

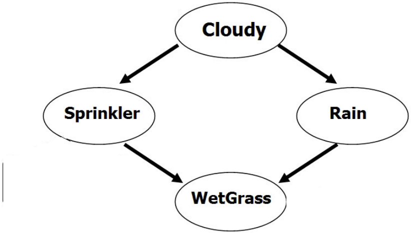


Figure: Is the grass wet?

Another Bayesian network

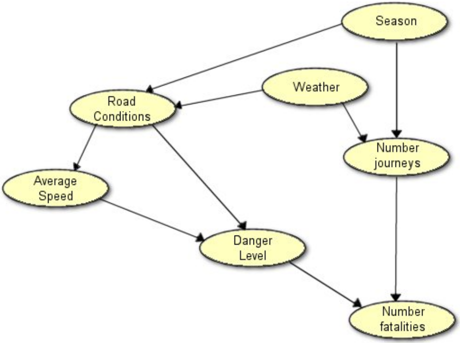
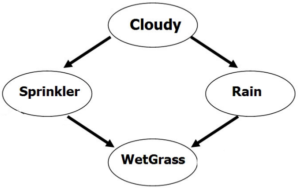


Figure: Traffic fatalities

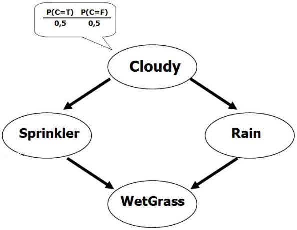
Explanation



- Nodes denote random events (here: Y/N)
- Arrows denote conditioning
- There is no rain $\rightarrow$ sprinkler arrow (but there could be one)
- This is a **directed acyclic graph (dag)**

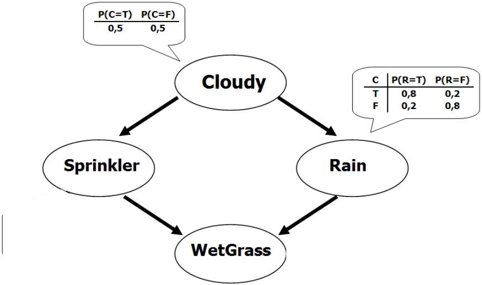
Specify probabilities

... at nodes without "in" arrows



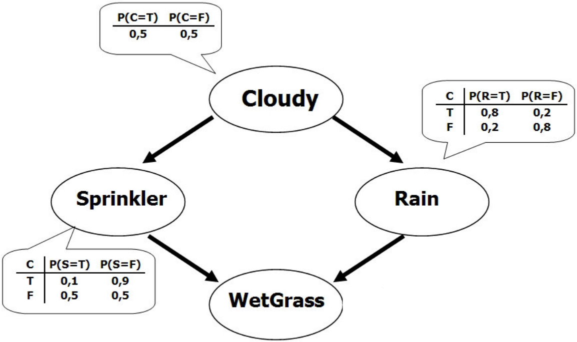
Specify conditional probabilities

... at other nodes.



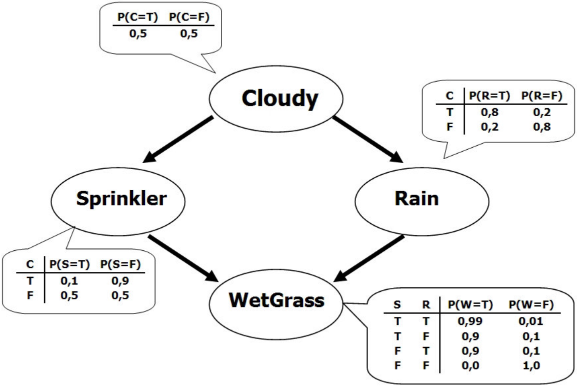
This means $\mathcal{P}(rain|cloudy) = .8$, $\mathcal{P}(rain| \sim cloudy) = .2$ *and so on. In practice, these have to be learned from data.*

Specify conditional probabilities



Row sums are 1, but column sums need not be.

Specify conditional probabilities



There are four possible conditioning events for wet grass. What is redundant in the table?

Compute some probabilities

Probability that the grass is wet

$$\begin{aligned}
\mathcal{P}(\mathit{cloudy}) &= .5, \quad \mathcal{P}(\sim \mathit{cloudy}) = .5 \\
\mathcal{P}(\mathit{sprinkler}) &= .1 \cdot .5 + .5 \cdot .5 = .3 \\
\mathcal{P}(\sim \mathit{sprinkler}) &= .7 \\
\mathcal{P}(\mathit{rain}) &= .8 \cdot .5 + .2 \cdot .5 = .5 = \mathcal{P}(\sim \mathit{rain}) \\
\mathcal{P}(\mathit{wet\ lawn}) &= .99 \cdot .3 \cdot .5 + .9 \cdot .3 \cdot .5 \\
&\quad + .9 \cdot .7 \cdot .5 + .0 \cdot .3 \cdot .5 \\
&\boxed{\approx .6}
\end{aligned}$$

Some inferences

Suppose the sprinkler was on.

$$
\begin{aligned}
\mathcal{P}(\textit{cloudy}) &= .5, \quad \mathcal{P}(\sim \textit{cloudy}) = .5 \\
\mathcal{P}(\textit{sprinkler}) &= 1 \\
\mathcal{P}(\sim \textit{sprinkler}) &= 0 \\
\mathcal{P}(\textit{rain}) &= .5 = \mathcal{P}(\sim \textit{rain}) \\
\mathcal{P}(\textit{wet lawn}) &= .99 \cdot 1 \cdot .5 + .9 \cdot 1 \cdot .5 \\
&\quad + .9 \cdot 0 \cdot .5 + .0 \cdot 0 \cdot .5 \\
&\boxed{\approx .945}
\end{aligned}
$$

More inferences

Suppose the sprinkler was on. What is the probability that it rained? *Need to reverse the conditioning.*

$$\mathcal{P}(cl|spr) = \mathcal{P}(spr|cl) \cdot \frac{\mathcal{P}(cl)}{\mathcal{P}(spr)} = .1 \cdot \frac{.5}{.7} \approx .07$$

We know that the sprinkler was on, so $\mathcal{P}(spr) = 1$ and $\mathcal{P}(cl) = .07$, $\mathcal{P}(\sim cl) = .93$. Then

$$\begin{aligned}\mathcal{P}(rain) &= \mathcal{P}(rain|cl) \cdot \mathcal{P}(cl) + \mathcal{P}(rain| \sim cl) \cdot \mathcal{P}(\sim cl) \\ &= .8 \cdot .07 + .2 \cdot .93 \approx \boxed{.24}\end{aligned}$$

Questions

- The lawn is wet. Find $\mathcal{P}(cloudy)$.
- What changes if the network looks like this: