9.5.2012

Today's topics:

1) More examples of distributions, including discrete random variables.
2) Conditional and joint densities.

Suggested reading:
sections 2.5–2.6

You requested more examples.
Let's discuss a few more random variables.

**Laplace distribution**

\[ f_X(x) = \frac{c}{2} e^{-c|x|}, \quad x \in \mathbb{R}, \quad c > 0. \]

Note that the Gaussian and Laplace distributions are both special cases of the generalized Gaussian distribution, also called exponential families:

\[ f_X(x) = c_1 e^{-c_2 |x|^c_3}. \]

The Laplace is often used to model DCT coefficients in image compression.

[Diagram of image (pixels) with arrows to discrete cosine transformation and DCT coefficients]
The DCT coefficients are encoded into bits by:
1) discretizing the continuous-valued coefficients to a discrete set of representation levels.
2) Representation levels are assigned probabilities based on Laplace distribution.
3) Levels that are more probable are encoded with fewer bits—like Morse code, which has short codes for common letters.

**Poisson distribution.**

Consider a digital camera. While the shutter is open, suppose that on average of 100 photons hit the photon detector. How do we compute the distribution of the number of photons?

To model this, keep in mind that at every instant a photon could arrive. Consider $n \to 1$ Bernoulli trials where the parameter $p$ is maintained such that $p \cdot n = 100$. Each trial coincides to a time period of duration $\frac{1}{n}$ the time that the shutter was open. As $n \to \infty$, it's very unlikely that more than one photon arrived during this time, so the Bernoulli model is reasonable.
What's the probability of $k$ photons?

$$Pr(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Note that $\frac{n!}{(n-k)! k!} \approx \frac{n^k}{k!}$ for large $n$, so

$$Pr(k) \approx \frac{n^k}{k!} \frac{\mu^k}{\mu^k} (1 - \frac{\mu}{n})^{n-k}$$

Note that as $n \to \infty$ we can use the Taylor series:

$$\left(1 + \frac{x}{n}\right)^n \approx e^x$$

$$Pr(k) \approx \frac{\mu^k}{k!} e^{-\mu}$$

We began with continuous-valued random variables (Gaussian, uniform, Laplace), and then discussed Poisson, which is a discrete random variable.

Instead of a probability distribution function $f_X(x)$ for discrete RV's we have a probability mass function (PMF), $P_X(x)$. 

Active learning exercise
(From last time)
A random student fills out random true/false answers to 5 questions.

(1) Compute an expression for \( P_x(c) \).

(2) Compute

\[
P_x(0) = \\
P_x(1) = \\
P_x(2) = \\
P_x(3) = \\
P_x(4) = \\
P_x(5) = \\
\]

(3) Sketch the CDF.
Can we express the PMF as a pdf? For a discrete RV, the CDF contains discontinuities. Seeing that the pdf is the derivative of the CDF, the CDF will contain delta functions.

Note - delta functions are described in Appendix 13.

Example (Bernoulli random variable.)

\[ N = \{H, T\}, \quad X(H) = 0, \quad X(T) = 1. \]

Suppose \[ P(H) = p, \quad P(T) = 1 - p. \]

The pdf has following form:

\[ f_X(x) = p \delta(x) + (1-p) \delta(x-1) \]
Geometric RV example

\[ P_X(k) = \begin{cases} \frac{p}{(1-p)^k} & 0 \leq k < \infty \\ 0 & \text{else} \end{cases} \]

where \( k \in \mathbb{Z} \) is integer.

The interpretation of a geometric RV can be understood in the context of buying a new car. With probability \( p \), the car we look at is good and we buy it; else we take a look at another car.

\( X = k \) means that we looked at \( k \) cars before seeing the one that we bought. This means that \( k \) times we dislike the car (with prob \( 1-p \) each time) and once we like the car (with prob \( p \)).

Note that

\[ \sum_{k=0}^{\infty} P_X(k) = \sum_{k=0}^{\infty} p \left(1-p\right)^k \]

\[ = p \cdot \frac{1}{1-(1-p)} = \frac{p}{p} = 1. \]
Conditional and joint distributions

When we studied Bayes’ law, we were interested in the probability of an event $E$ conditioned on event $F$ happening (or not). Many times we’re interested in the distribution of a RV conditioned on an event, where the event could be some RV taking on some value or range of values.

Examples
1) distribution of number of photons conditioned on at least 50 photons arriving.
2) distribution of snowfall conditioned on temperature above 30°F.

These can be written as a CDF,

$F_X(x | B) = \frac{\Pr(X \leq x) | B)}{\Pr(B)}$

Note that we divided the joint probability by the probability of the conditioning event, as before.

Example

$\Pr(k \text{ photons } | X \geq 50) = \frac{\Pr(X \leq k, X \geq 50)}{\Pr(X \geq 50)}$
Joint cumulative distribution function

Let $X_1, ..., X_n$ be RV's over $(\mathbb{R}, \mathcal{F}, P)$.

$F_{X_1, ..., X_n}(x_1, ..., x_n) = P(X_1 \leq x_1, ..., X_n \leq x_n)$

**Example (n=2) via active learning**

Want to compute $Pr(\text{rectangle})$

To keep it simple, suppose that probabilities of borders (lines) are zero.

Graphically, this can be seen as follows.

$Pr(\text{(1)}) = F(a_1, b_1)$

$Pr(\text{(1)}) + Pr(\text{(2)}) = F(a_1, b_2)$

$Pr(\text{(1)}) + Pr(\text{(3)}) = F(a_2, b_1)$

$Pr(\text{(1)}) + Pr(\text{(2)}) + Pr(\text{(3)}) + Pr(\text{(4)}) = F(a_2, b_2)$
(1) Compute $Pr\left(\mathcal{G}_2\right)$ using $F_{X_1,X_2}(a_1,b_1)$ and $F_{X_1,X_2}(a_2,b_2)$.

(2) Compute $Pr\left(\mathcal{G}_3\right)$ using $F_{X_1,X_2}(a_1,b_1)$ and $F_{X_1,X_2}(a_2,b_1)$.

(3) Compute $Pr(\text{rectangle}) = Pr\left(\mathcal{G}_4\right)$ using $F_{X_1,X_2}(a_1,b_1)$, $F_{X_1,X_2}(a_1,b_2)$, $F_{X_1,X_2}(a_2,b_1)$ and $F_{X_1,X_2}(a_2,b_2)$. 
Joint probability density function
The joint pdf and joint CDF relate as follows:

\[ F_{x_1, \ldots, x_n}(x_1, \ldots, x_n) = \int_{z_{1}}^{x_1} \ldots \int_{z_{n}}^{x_n} f_{x_1, \ldots, x_n}(z_1, \ldots, z_n) \, dz_1 \ldots dz_n \]

Marginalization Given joint CDF \( F_{x_1, x_2} \), can compute \( F_{x_1}(x_1) = \Pr(X_1 \leq x_1) \)

\[ \lim_{x_2 \to \infty} F_{x_1, x_2}(x_1, +\infty) \]

notation \( \rightarrow = F_{x_1, x_2}(x_1, +\infty) \)

\[ = \int_{-\infty}^{x_1} \int_{-\infty}^{+\infty} f_{x_1, x_2}(u_1, u_2) \, du_2 \, du_1 \]

Example

<table>
<thead>
<tr>
<th>( x_1 = 0 )</th>
<th>( x_1 = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_2 = 0 )</td>
<td>0.1</td>
</tr>
<tr>
<td>( x_2 = 1 )</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\( \Pr(X_1 \leq 0) = 0.1 + 0.3 = 0.4 \)

Independent random variables

\( F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y) \)

\( f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) \)

\[ = \frac{\partial}{\partial x} F_X(x) \frac{\partial}{\partial y} F_Y(y) = f_X(x) \cdot f_Y(y) \]

Therefore, irrespective of \( y \) we get

\( f_X(x \mid Y \leq y) = f_X(x) \)