Lecture 13: limits of probabilities of nested events Kelly 2.5

13.1 The collection of all events

For (finte or) countable Ω, events are all subsets of Ω, but this does not work for $\Omega = \Re$.

More generally, Ω is event, E an event $\Rightarrow E^c$ an event, and $E_1, E_2, ...$ events $\Rightarrow \bigcup_{i=1}^{\infty}$ an event.

(Such collections, closed under complements and countable unions, are called σ-fields: Kelly 1.3.)

Note $\Phi = \Omega^c$ is an event, and $\bigcap E_i = (\bigcup E_i^c)^c$ are then also events.

13.2 The events on $\Re$ generated by sets $(-\infty, b]$ for all real b.

(i) First, for $a < b$, $(-\infty, a]^c \bigcap (-\infty, b] = (a, b]$ so we have the half-left-open intervals.

(ii) Next, $\bigcup_n (a, b - \frac{1}{n}] = (a, b)$ so we have the open intervals.

(iii) Next, $\bigcap_n (a - \frac{1}{n}, b] = [a, b]$ so we have the closed intervals.

(iv) Next, if $a = b$ in (iii), we have isolated points $\{b\}$.

(v) Finally, taking countable unions of these, we have all countable sets.

Sets generated by countable unions and intersections of intervals, are the *Borel sets* (Kelly 1.3), and make up the set of events for a real-valued (discrete or continuous) random variable.

13.3 Increasing and decreasing nested sets

(i) $A_1,\ A_2,\ , A_3, ...$ are nested increasing sets if $A_1 \subset A_2 \subset A_3 \subset .....$ Then $\bigcup_1^n A_i = A_n$ and $\bigcap_1^n A_i = A_1$.

(ii) $A_1,\ A_2,\ , A_3, ...$ are nested decreasing sets if $A_1 \supset A_2 \supset A_3 \supset .....$ Then $\bigcup_1^n A_i = A_1$ and $\bigcap_1^n A_i = A_n$.

Examples: Kelly 2.5.1, 2.5.2

1. Increasing: $A_n = (-\infty, x - \frac{1}{n}]$. $\lim_{n\to\infty} A_n = (-\infty, x)$.
2. Decreasing: $A_n = (-\infty, x + \frac{1}{n})$ or $A_n = (-\infty, x + \frac{1}{n}]$. $\lim_{n\to\infty} A_n = (-\infty, x]$.
3. Decreasing: A_n is event of no successes in n tries. Then $\lim_{n\to\infty} A_n$ is event of no success ever.

13.4 Nested sets theorem; Kelly 2.5.3, 2.5.8

Theorem: Let $A_1, A_2,\ ....$ be any events in Ω.

(i) If $A_1 \subset A_2 \subset A_3 \subset ....$, $P(A_1 \bigcup A_2 \bigcup A_3.....) = \lim_{n\to\infty} P(A_n)$.

(ii) If $A_1 \supset A_2 \supset A_3 \supset ....$, $P(A_1 \bigcap A_2 \bigcap A_3.....) = \lim_{n\to\infty} P(A_n)$.

Proof: (i) Let $B_i = A_i \cap A_{i-1}^c$; Then B_i are disjoint and $B_1 \bigcup B_2 \bigcup ... \bigcup B_n = A_1 \bigcup A_2 \bigcup ... \bigcup A_n$, so
$P(A_1 \bigcup A_2 \bigcup A_3.....) = P(B_1 \bigcup B_2 \bigcup B_3.....) = \sum_1^{\infty} P(B_i) = \lim_{n\to\infty}(\sum_1^n P(B_i)$
$= \lim_{n\to\infty} P(B_1 \bigcup B_2 \bigcup .... \bigcup B_n) = \lim_{n\to\infty} P(A_n)$.

(ii) Let $D_i = A_i^c$, so from (i) $P(D_1 \bigcup D_2 \bigcup D_3.....) = \lim_{n\to\infty} P(D_n)$.
But $P(D_1 \bigcup D_2 \bigcup D_3.....) = P((A_1 \bigcap A_2 \bigcap ...)^c) = 1 - P(A_1 \bigcap A_2 \bigcap ....)$
and $\lim_{n\to\infty} P(D_n) = \lim_{n\to\infty}(1 - P(A_n)) = 1 - \lim_{n\to\infty} P(A_n)$.

13.5 Examples; Kelly 2.5.4 etc.

(i) In independent trials, with probability of success $p > 0$, eventually we have success with probability 1, since if A_n is event of no successes in n tries, $P(A_n) = (1-p)^n \longrightarrow 0$.

(ii) Let the probability of success on try k be p_k. Let $D_n = A_n^c$ be event of success by try n. Then $P(D_n) \leq \sum_{k=1}^n p_k$. If p_k decrease fast (e.g. $p_k = 0.1/n^2$) then $\lim P(D_n) < 1$; eventual success is not certain.

(iii) Increasing: For any random variable X; $P(X < a) = \lim P(X \leq (a - \frac{1}{n}))$.

Increasing: For any random variable X; $P(X > a) = \lim P(X > (a + \frac{1}{n}))$.

Decreasing: For any random variable X; $P(X \leq a) = \lim P(X \leq (a + \frac{1}{n}) = \lim P(X < (a + \frac{1}{n}))$.

Decreasing: For any random variable X; $P(X \geq a) = \lim P(X \geq (a - \frac{1}{n}) = \lim P(X > (a - \frac{1}{n}))$.

Lecture 14: Cumulative distribution functions Kelly 3.2

14.1 (i) Definition: For any random variable X, the *cumulative distribution function* is defined as

$$F_X(x) \;=\; P(X \leq x) \text{ for } -\infty < x < \infty.$$

(ii) For a discrete random variable with pmf $p_X(x)$, $F_X(b) \;=\; \sum_{x \leq b} p_X(x)$.

(iii) For a continuous random variable with pdf $f_X(x)$, $F_X(b) \;=\; \int_{-\infty}^{b} f_X(x)dx$.

(iv) For all random variables, $P(a < X \leq b) = F(b) - F(a)$
because $\{X \leq b\} \;=\; \{X \leq a\} \cup \{a < X \leq b\}$ and $\{X \leq a\} \cap \{a < X \leq b\} \;=\; \Phi$ (empty set).

14.2 Properties:

(i) F_X is a non-decreasing function: if $a < b$, then $F_X(a) \leq F_X(b)$, because $\{X \leq a\} \subset \{X \leq b\}$.

(ii) $\lim_{b \to \infty} F_X(b) \;=\; 1$, because for any increasing sequence $b_n \to \infty$, $n = 1, 2, 3, ...,$
$\Omega = \{X < \infty\} = \cup\{X \leq b_n\}$, so $1 = P(\Omega) \;=\; \lim_{n \to \infty} P(X \leq b_n) \;=\; \lim_{n \to \infty} F_X(b_n)$.

(iii) $\lim_{b \to -\infty} F_X(b) \;=\; 0$, because for any decreasing sequence $b_n \to -\infty$, $n = 1, 2, 3, ...,$
$\Phi = \{X = -\infty\} = \cap\{X \leq b_n\}$, so $0 = P(\Phi) \;=\; \lim_{n \to \infty} P(X \leq b_n) \;=\; \lim_{n \to \infty} F_X(b_n)$.

(iv) F_X is right-continuous. That is, for any b and any decreasing sequence b_n, $n = 1, 2, 3, ...,$ with $b_n \to b$ as $n \to \infty$, $\lim_{n \to \infty} F_X(b_n) = F_X(b)$, because $\{X \leq b\} \;=\; \cap\{X \leq b_n\}$.

Note $P(X \leq b) \;=\; P(X < b) \;+\; P(X = b)$, and $P(X < b) \;=\; \lim_{x \to b^-} F(x)$.
If X is discrete, with $P(X = b) > 0$, F_X will be discontinuous at $x = b$.

14.3 Case of continuous random variables:

For discrete random variables, $F_X(x)$ is just a set of flat (constant) pieces, with jumps in amount $P(X = x_i)$ at each possible value x_i of X. This is not very useful.

For continuous random variables, the cdf is very useful!

$$F_X(x) \;=\; P(X \leq x) \;=\; \int_{-\infty}^{x} f_X(w)dw \quad \text{so} \quad \frac{dF_X(x)}{dx} \;=\; f_X(x).$$

That is, we get the pdf by differentiating the cdf: the cdf is often easier to consider.

Example: scaling an exponential random variable.

Suppose $f_X(x) \;=\; \lambda e^{-\lambda x}$ on $x \geq 0$, and let $Y = aX$ $(a > 0)$. What is the pdf of Y?

$$\begin{aligned}
\text{First, } F_x(x) &= \int_0^x \lambda e^{-\lambda w} dw \;=\; [-e^{-\lambda w}]_0^x \;=\; 1 - e^{-\lambda x} \text{ on } x \geq 0. \\
\text{Now, } F_Y(y) &= P(Y \leq y) \;=\; P(aX \leq y) \;=\; P(X \leq y/a) \;=\; F_X(y/a) \;=\; (1 - e^{\lambda y/a}), \\
\text{so } f_Y(y) &= F_Y'(y) \;=\; \frac{d}{dy}(1 - e^{-\lambda y/a}) \;=\; (\lambda/a)e^{-(\lambda/a)y} \text{ on } y \geq 0.
\end{aligned}$$

That is Y is an exponential random variable with parameter λ/a.

14.4 Using the cdf to consider functions of random variables (Kelly 5.2)

Using the cdf is often the easiest way to consider functions of a random variable.

Example: Suppose X is Uniform U(0,1). What is the pdf of $Y = X^3$?

$$\begin{aligned}
f_X(x) &= 1,\; 0 \leq x \leq 1; \quad F_X(x) \;=\; x,\; 0 \leq x \leq 1 \\
F_Y(y) \;=\; P(Y \leq y) \;=\; P(X^3 \leq y) &= P(X \leq y^{1/3}) \;=\; F_X(y^{1/3}) \;=\; y^{1/3},\; 0 \leq y \leq 1 \\
f_Y(y) &= \frac{d}{dy}F_Y(y) \;=\; (1/3)y^{-2/3}\; 0 \leq y \leq 1
\end{aligned}$$

Lecture 15: Conditional probability for random variables

15.1 Conditioning a discrete random variable

Recall $P(X=x)$ is just an event, and $P(X \in B) = \sum_{x \in B} P(X=x)$.

So $P(X \in C \mid X \in B) \;=\; P(X \in B \bigcap C)/P(X \in B)$.

Also $P(X=x \mid X \in B) \;=\; P(X=x)/P(X \in B)$, provided $x \in B$.

15.2 Examples of conditioning a discrete random variable

(i) If X is Poisson, parameter 1: $P(X=x) = e^{-1}1^x/x! \;=\; e^{-1}/x!$, $P(X \geq 2) \;=\; 1 - e^{-1} - e^{-1}$, and $P(X=x \mid X \geq 2) \;=\; (e^{-1}/x!)/(1-2e^{-1})$, for $x = 2, 3, 4, ...$

(ii) If X is Bin(11,0.5): $P(X \text{ even}) \;=\; 1/2$ and $P(X=x \mid X \text{ even}) \;=\; 2P(X=x)$ if x is even, 0 otherwise.

(iii) **The forgetting property of the Geometric Distribution**

Suppose on each try the probability of "success" is p. Let X be the number of failures before a "success" is achieved. Then $P(X=x) \;=\; (1-p)^x.p$ for $x = 0, 1, 2, 3, ...$ (Geometric distribution).

$P(X \geq k) \;=\; P(\text{first } k \text{ are failures}) \;=\; (1-p)^k$ (or $P(X \geq k) \;=\; \sum_{x=k}^{\infty}(1-p)^x.p \;=\;$ same thing).

$P(X \geq k+\ell \mid X \geq \ell) \;=\; P(X \geq k+\ell)/P(X \geq \ell) \;=\; (1-p)^{k+\ell}/(1-p)^{\ell} \;=\; (1-p)^k \;=\; P(X \geq k)$.

15.3 Conditioning a continuous random variable

Recall $X \in B$ is an event and $P(X \in B) = \int_{x \in B} f_X(x)dx$.

So $P(X \in C | X \in B) \;=\; P(X \in B \cap C)/P(X \in B) \;=\; \int_{B \cap C} f(x)dx / \int_B f(x)dx$

15.4 Examples of conditioning a continuous random variable

(i) Example of a Uniform random variable Suppose X has p.d.f. $f(x) = 1$, $0 \leq x \leq 1$.

$P(X > 0.6 \mid X \leq 0.8) \;=\; P(0.6 < X \leq 0.8)/P(X \leq 0.8) \;=\; 0.2/0.8 \;=\; 0.25$.

(ii) Example for an exponential random variable

Suppose X has p.d.f. $f(x) = 0.5e^{-0.5x}$ on $0 < x < \infty$: $F(x) \;=\; \int^x f(w)dw \;=\; 1 - e^{-0.5x}$.

So $P(X \leq 6 \mid X > 2) = P(2 < X \leq 6)/P(X > 2) \;=\; (e^{-1} - e^{-3})/e^{-1} \;=\; (1 - e^{-2}) \;\approx\; 6/7$.

(iii) The **forgetting property** of the exponential.

Suppose X has p.d.f. $f(x) = \lambda \exp(-\lambda x)$, $\quad 0 < x < \infty$.

Note $F_X(a) = P(X \leq a) \;=\; \int_0^a f(x)\,dx \;=\; (1 - \exp(-\lambda a))$, so $P(X > a) \;=\; \exp(-\lambda a)$. Consider $P(X > a+b \mid X > a) = P(X > a+b)/P(X > a) = \exp(-\lambda(a+b))/\exp(-\lambda a) = \exp(-\lambda b) = P(X > b)$.

15.5 Approximating discrete by continuous distributions

Note the geometric and exponential distributions both have the "forgetting" property.

In fact, geometric is just a discrete version of the exponential.

Consider a r.v. T with geometric distribution with $p = 0.02$ and also T with an exponential with rate parameter $\lambda = 0.02$.

(i)Compare $P(T \geq 50)$ under the two models.

(ii) Repeat for $p = 0.001$ (geometric) and $\lambda = 0.001$ (exponential).

(iii) Using either model (p or $\lambda = 0.02$), what is $P(T \geq 100 \mid T \geq 50)$?

(iv) Using either model (p or $\lambda = 0.02$), what is $P(T \leq 50 \mid T \leq 100)$?

(v) In (i) is it better to compare $P(T > 49.5)$? Why? (Actually not, but it could be?)