STATISTICS 205A Spring 1999. *David Aldous.*
Lecture 1.
(i) Constructing random variables.
(ii) Radon-Nikodym densities.

A r.v. X with values in a measurable space $(S, \mathcal{S})$ has a <u>distribution</u> ν:

$$\nu(A) = P(X \in A), A \in \mathcal{S}.$$

Question: given a p.m. ν, does there exist a r.v. X whose distribution is ν? Uninteresting answer: Yes, because we can take $\Omega = S$ and $X =$ identity.

To get something more interesting, recall undergraduate result.

Lemma 1 *Let μ be a probability measure on R, let $F(x) = \mu(-\infty, x]$ be its distribution function, let*

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \ 0 \leq u \leq 1$$

be the inverse distribution function. Then

$$F^{-1}(U) \text{ has distribution } \mu$$

where U has $U(0,1)$ distribution.

Now consider S-valued r.v.'s of the form $h(U)$, where $h : [0,1] \to S$ is measurable.

Lemma 2 *Let ν be a p.m. on a <u>nice</u> (= <u>Standard Borel</u>: p. 33) space. Then there exists measurable $h : [0,1] \to S$ such that $h(U)$ has distribution ν.*

Proof. Easy: use Lemma 1 and definition of nice: there exists $1-1$ map $\phi : S \to R$ with ϕ and ϕ^{-1} measurable.

To apply we need (Theorem 1.4.12): any complete separable metric space is nice.

Corollary 3 *(Counter-intuitive?). Let $X_1, X_2, \ldots$ be R-valued. Then there exist measurable $h_1, h_2, \ldots$ such that $(h_1(U), h_2(U), \ldots)$ has the same (joint) distribution as $(X_1, X_2, \ldots)$.*

Proof. Use idea: consider $\mathbf{X} = (X_1, X_2, \ldots)$ as a single R^∞-valued r.v.

Here's a more constructive approach. Consider the binary representation of reals in $(0,1)$

$$U = \sum_{i=1}^{\infty} B_i 2^{-i}.$$

The B's are independent Bernoulli $(1/2)$. For each $k \geq 1$ let $I^{(k)} = (i_{k1}, i_{k2}, \ldots)$ be an infinite sequence of integers, the sequences disjoint in k. Use the B's from $I^{(k)}$ to define U_k:

$$U_k = \sum_{j=1}^{\infty} B_{i_{kj}} 2^{-j}.$$

Then the U's are independent $U(0,1)$. Apply Lemma 1:

Corollary 4 *Let $\theta_1, \theta_2, \ldots$ be p.m.'s on R. Then there exist independent r.v.'s $X_1, X_2, \ldots$ such that X_i has distribution θ_i for each i.*

Note this does not use Kolmogorov extension – later we will give a "constructive" proof of the Kolmogorov extension theorem.

Radon-Nikodym densities.

If you haven't seen this stuff in a measure theory course, read Appendix 8 and try the exercises.

Lecture 2.

Want to formalize the idea "conditional distribution of X_2 given $X_1 = s_1$. We could write

$$Q(s_1, B) = P(X_2 \in B | X_1 = s_1).$$

What sort of object is Q?

Measure-theory set-up. $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ are measure spaces, and $(S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$ is their product space. A <u>kernel Q from S_1 to S_2</u> is a map $Q : S_1 \times \mathcal{S}_2 \to R$ such that

(a) $B \to Q(s_1, B)$ is a p.m. on $(S_2, \mathcal{S}_2)$ for each fixed $s_1 \in S_1$

(b) $s_1 \to Q(s_1, B)$ is a measurable function $S_1 \to R$ for each fixed $B \in \mathcal{S}_2$.

If S_1 and S_2 are countable then kernels correspond to <u>stochastic matrices</u>.

In undergraduate course, continuous r.v.'s (X, Y) have a joint density $f(x, y)$, a marginal density $f(x)$ for X, and a conditional density $f(y|x)$ for Y given $X = x$: these are related by

$$f(x, y) = f(x) f(y|x).$$

Proposition 5 *Given a p.m. μ on $S_1 \times S_2$, a p.m. μ_1 on S_1 and a kernel Q from S_1 to S_2, the following are equivalent.*

$$\mu(A \times B) = \int_A Q(s, B) \mu_1(ds); \;\; A \in \mathcal{S}_1, B \in \mathcal{S}_2. \tag{1}$$

$$\mu(D) = \int_{S_1} Q(s_1, D_{s_1}) \mu(ds_1); \;\; D \in \mathcal{S}_1 \times \mathcal{S}_2 \tag{2}$$

where $D_{s_1} = \{s_2 : (s_1, s_2) \in D\}$.

$$\int_{S_1 \times S_2} h(s_1, s_2) \mu(d\mathbf{s}) = \int_{S_1} \left(\int_{S_2} h(s_1, s_2) Q(s_1, ds_2) \right) \mu_1(ds_1) \tag{3}$$

for all measurable $h : S_1 \times S_2 \to R$ for which either $h \geq 0$ or h is μ-integrable.

Note: part of assertion of (2,3) is that integrands are measurable.

Jargon: I call Q the <u>conditional probability kernel</u> for μ, but this isn't standard.

Lemma 6 *For each $D \in \mathcal{S}_1 \times \mathcal{S}_2$*

(i) $D_{s_1} \in \mathcal{S}_2$ for all $s_1 \in S_1$

(ii) the map $s_1 \to Q(s_1, D_{s_1})$ is measurable.

Proof. Apply $\pi - \lambda$ theorem (1.4.2) to class $\mathcal{D}$ of sets D for which assertions are true.

Proof of Proposition 5. (1) $\rightarrow$ (2). Lemma 6 says (2) is meaningful: consider class of D's where it is true. True for $D = A \times B$ by (1). Apply $\pi - \lambda$ theorem.

(2) $\rightarrow$ (3). Conclusion is meaningful and true for $h = 1_D$, and hence for simple h. General $h \geq 0$ is increasing limit of simple h_n defined by

$$h_n(\cdot) = \min(n, 2^{-n}\lfloor h(\cdot)2^n \rfloor)$$

so by monotone convergence, result holds for $h \geq 0$. For general h write $h = h^+ - h^-$.

Theorem 7 *[easy part] Let μ_1 be a p.m. on $\mathcal{S}_1$ and let Q be a kernel from S_1 to S_2. Then there exists a unique p.m. μ on $S_1 \times S_2$ such that the relations of Proposition 5 hold.*

Conversely, let μ be a p.m. on $S_1 \times S_2$. Define μ_1 by: $\mu_1(A) = \mu(A \times S_2)$. Then [hard part: 4.1.6] provided S_2 is nice, there exists a kernel Q from S_1 to S_2 such that the relations of Proposition 5 hold.

Proof. [easy part] Use (2) to define $\mu(D)$: this makes sense because of Lemma 6. Need to verify μ is a p.m. Issue is countable additivity. If $D^n \uparrow D$ then $D^n_{s_1} \uparrow D_{s_1}$, so $Q(s_1, D^n_{s_1}) \uparrow Q(s_1, D_{s_1})$, so $\mu(D^n) \uparrow \mu(D)$.

[hard part] As with Lemma 2 we can reduce to the case $S_2 = R$. Write $S_1 = S$. Let r denote a rational. We shall use easy analysis fact. Let $F(r)$ be a real-valued function defined on the rationals and such that

$$F(r) \text{ is non-decreasing.} \tag{4}$$

$$F \text{ is right-continuous on rationals} \tag{5}$$

$$\lim_{r \to -\infty} F(r) = 0, \lim_{r \to \infty} F(r) = 1. \tag{6}$$

Then F extends to a distribution function, by setting

$$F(x) = \lim_{r \downarrow x} F(r).$$

For each r let ν_r be the (sub-probability) measure on S defined by

$$\nu_r(A) = \mu(A \times (-\infty, r]).$$

So $\nu_r(A) \leq \mu_1(A)$. Let $F(s, r)$ be the Radon-Nikodym density of ν_r with respect to μ_1. That is to say

$$s \to F(s, r) \text{ is measurable}$$

$$\mu(A \times (-\infty, r])) = \int_A F(s,r)\mu_1(ds) \text{ for all } A.$$

We now modify F on μ_1-null sets so that, for each s, the maps $r \to F(s,r)$ will satisfy (4 - 6). For $r_1 < r_2$,

$$\int_A (F(s,r_2) - F(s,r_1))\mu_1(ds) = \mu(A \times (r_1, r_2]) \geq 0 \text{ for all } A$$

and so the integrand is a.e. non-negative. Modify to make it everywhere non-negative. Similarly, consider $r_n \downarrow r$. Then $\mu(A \times (r, r_n]) \downarrow 0$ and so $F(s, r_n) \downarrow F(s,r)$ μ_1-a.e., and the null set depends only on r. So we can modify to make $F(s, \cdot)$ right-continuous on rationals, for all s. Finally, easy to modify to get

$$\lim_{r \to -\infty} F(s,r) = 0, \quad \lim_{r \to \infty} F(s,r) = 1 \text{ for all } s.$$

So by analysis fact, $F(s, \cdot)$ extends to a distribution function. Define $Q(s, \cdot)$ to be the p.m. whose distribution function is $F(s, \cdot)$. To finish the proof, we must show: for each $B \subset R$

$$s \to Q(s,B) \text{ is measurable}$$

$$\mu(A \times B) = \int_A Q(s,B)\mu_1(ds); \text{ all } A \subset S.$$

By construction these hold for $B = (-\infty, r]$. Apply the $\pi - \lambda$ theorem.

Lecture 3.

Topics: Uses of Fubini's theorem, Kolmogorov extension theorem.

Given p.m.'s μ_1 on S_1 and μ_2 on S_2 we can define the product measure $\mu = \mu_1 \times \mu_2$ on $S_1 \times S_2$, which has properties (7 - 9) below. These properties follow from Theorem 7, putting $Q(s_1, \cdot) = \mu_2(\cdot)$.

$$\mu(A \times B) = \mu_1(A)\mu_2(B); A \subset S_1, B \subset S_2 \tag{7}$$

$$\mu(D) = \int_{S_1} \mu_2(D_{s_1})\ \mu_1(ds_1);\ D \subset S_1 \times S_2 \tag{8}$$

For measurable $h : S_1 \times S_2 \to R$ with either $h \geq 0$ or h is μ-integrable,

$$\int_{S_1 \times S_2} h(\mathbf{s})\mu(d\mathbf{s}) = \int_{S_1} \left(\int_{S_2} h(s_1, s_2)\mu_2(ds_2) \right)\ \mu_1(ds_1) \tag{9}$$

$$= \int_{S_2} \left(\int_{S_1} h(s_1, s_2)\mu_1(ds_1) \right)\ \mu_2(ds_2)$$

The final equalities are Fubini's Theorem. These results also hold for σ-finite measures. See Appendix 6 for examples illustrating the necessity of the hypotheses. Here are some more "practical" examples. Here X, Y denote real-valued r.v.'s with distributions μ, ν, and λ is Lebesgue measure on the line.

Example. If $X \geq 0$ then $EX = \int_0^\infty P(X > t)dt$.

Proof. Apply Fubini's theorem to the set $D = \{(x,t) : x \geq t\} \subset [0, \infty) \times [0, \infty)$ and the product measure $\mu \times \lambda$.

Example. Parseval's identity. Let X have characteristic function $\phi(t) = E\exp(itX)$ and Y have characteristic function $\hat{\phi}(t)$. Then $\int \phi(t)\nu(dt) = \int \hat{\phi}(t)\mu(dt)$.

Proof. Compute $E\exp(iXY)$.

Example. Suppose X and Y are independent, and set $S = X + Y$. In undergraduate course we see the convolution formula for densities:

$$f_S(s) = \int f_Y(s-x)f_X(x)dx$$

which assumes densities f_Y and f_X exist. A completely general version can be stated in terms of distribution functions as

$$F_S(s) = \int F_Y(s-x)\mu(dx).$$

In the case where Y does have a density f_Y

$$f_S(s) = \int f_Y(s-x)\mu(dx)$$

Example. Conditional densities. We used these to motivate kernels; now we can prove the following. Suppose (X,Y) has joint density $f(x,y)$. Define $f(y|x) = f(x,y)/f_X(x)$ where $f_X(x) > 0$. Define $Q(x,\cdot)$ to be the distribution with density $f(\cdot|x)$. Then Q is the conditional probability kernel for Y given X.

Proof. Use Fubini's theorem to verify (1):

$$P(X \in A, Y \in B) = \int_A Q(x,B)\mu(dx).$$

I will give the "probabilistic" proof of the (countable) Kolmogorov extension theorem. Appendix 7 gives the measure theory proof. Some texts give a version for uncountable families, but this has no practical use.

We start with a "random variable" version of Theorem 7.

Corollary 8 *Let (X,U) be independent r.v.'s such that U is uniform on $[0,1]$, and X takes values in S and has distribution μ_1. Let μ be a p.m. on $S \times R$ with marginal μ_1. Then there exists measurable $f : S \times [0,1] \to R$ such that*

$$\mu = \textit{dist}(X,Y), \quad \textit{for } Y = f(X,U).$$

Proof. Let Q be the conditional probability kernel from S to R associated with μ (Theorem 7). For each $x \in S$ let $f(x,\cdot)$ be the inverse distribution function for the p.m. $Q(x,\cdot)$. Lemma 1 says $f(x,U)$ has distribution $Q(x,\cdot)$. In terms of measures, this is:

$$\lambda\{u : f(x,u) \in B\} = Q(x,B),\ \ B \subset R.$$

We have to verify: for $A \subset S, B \subset R$

$$P(X \in A, Y \in B) = \mu(A \times B).$$

Easy.

Theorem 9 (Kolmogorov extension) *Let* $(\mu_n; 1 \le n < \infty)$ *be p.m.'s on* R^n*. Suppose they are* <u>*consistent*</u> *in the following sense. For each n, regard* μ_{n+1} *as a measure on* $R^n \times R$*: then the marginal of* μ_{n+1} *is* μ_n*. Then there exists a unique p.m.* μ_∞ *on* R^∞ *such that, writing* $R^\infty = R^n \times R^\infty$*, the marginal of* μ_∞ *is* μ_n*.*

Proof. Let $(U_1, U_2, \ldots)$ be independent $U(0,1)$, which exist by Corollary 4. Define $X_1 = F_{\mu_1}^{-1}(U_1)$. Inductively, suppose we have defined $\mathbf{X}_n = (X_1, \ldots, X_n)$ as a measurable function of $(U_1, \ldots, U_n)$ so that $\mathrm{dist}(\mathbf{X}_n) = \mu_n$. We shall define $\mathbf{X}_{n+1}$ as a measurable function of $(\mathbf{X}_n, U_{n+1})$. Then the induction goes through, and we can define a infinite sequence of r.v.'s $(X_n; 1 \le n < \infty)$. Clearly $\mu_\infty = \mathrm{dist}(X_n; 1 \le n < \infty)$ satisfies the conclusion of the Theorem.

To do the inductive step, just apply Corollary 8 with $X = \mathbf{X}_n$, $U = U_{n+1}$ and $\mu = \mu_{n+1}$ regarded as a measure on $R^n \times R$.

Lecture 4.

Conditional expectation. Read section 4.1.

Lecture 5.

Topics. Conditional expectations, conditional probabilities and regular conditional distributions (r.c.d.'s). Conditioning and independence. Conditional independence (see homework for definition).

Let's record two lemmas.

Lemma 10 *If $E(X|\mathcal{G})$ is a.s. equal to some $\mathcal{D}$-measurable r.v., and if $\mathcal{D} \subset \mathcal{G}$, then $E(X|\mathcal{D}) = E(X|\mathcal{G})$.*

Lemma 11 *If X and Y are conditionally independent given $\mathcal{G}$, and if V is $\mathcal{G}$-measurable, then X and (Y, V) are conditionally independent given $\mathcal{G}$.*

Also record basic property of r.c.d.'s. If Q is a r.c.d. for Z given U then

$$E(h(Z)|U)(\omega) = \int h(z)Q(\omega, dz).$$

Lecture 6. Measure-theory set-up for Markov chains.

This material is presented somewhat differently in Durrett 5.1 and 5.2. I want to emphasize the conditional independence aspects. The first result (I call it the splice lemma) gives the "conditionally independent" analog of product measure.

Lemma 12 *Let S_1, S_2, S_3 be nice spaces. Let μ_{12} be a p.m. on $S_1 \times S_2$ and μ_{23} be a p.m. on $S_2 \times S_3$ such that the marginals on S_2 coincide. Then there exists a unique probability measure μ on $S_1 \times S_2 \times S_3$ such that, writing $\mu = dist(X_1, X_2, X_3)$,*

(i) $dist(X_1, X_2) = \mu_{12}$ and $dist(X_2, X_3) = \mu_{23}$

(ii) X_1 and X_3 are conditionally independent given X_2.

Proof. We can specify μ on $S_1 \times S_2 \times S_3$ by specifying a marginal p.m. on $S_1 \times S_2$ and a kernel Q from $S_1 \times S_2$ to S_3. So let the marginal be μ_{12} and let the kernel be

$$Q((s_1, s_2), \cdot) = Q_{23}(s_2, \cdot)$$

where Q_{23} is the kernel from S_2 to S_3 associated with μ_{23}. Property (i) is easy. For (ii),

$$E(h(X_3)|X_1, X_2) = \int h(x)Q((X_1, X_2), dx)$$