STSCI 4780
Conditional distributions & Gibbs sampling

Tom Loredo, CCAPS & DSS, Cornell University

© 2018-04-13

Agenda

1 **Joint from conditionals**

2 Gibbs sampling

Joint distribution from conditionals?

The symmetric parameterization of the BVN has 5 parameters:

- Marginal means: μ_x, μ_y
- Marginal standard deviations: σ_x, σ_y
- Correlation coefficient: ρ

If we fix $(\mu_x, \sigma_x, \mu_y, \sigma_y)$ and vary ρ, we generate a family of distributions with *identical marginals but different joint distributions*

Specifying marginals does **not** *uniquely determine the joint*

Specifying one marginal and its associated conditional does give the joint:

$$
\begin{aligned}
p(x, y) &= p(x)\, p(y|x) \\
&= p(y)\, p(x|y)
\end{aligned}
$$

What about *specifying the two conditionals*?

Hammersly-Clifford theorem

We'll be evaluating joint, marginal, and conditional distributions for multiple choices of (x, y), so we introduce notation distinguishing the various functions (instead of using $p()$ for everything):

$$
\begin{aligned}
f(x, y) &\equiv p(x, y) \\
m_1(x) &\equiv p(x) = \int \mathrm{d}y \, p(x, y) \\
m_2(y) &\equiv p(y) = \int \mathrm{d}x \, p(x, y) \\
c_{12}(x; y) &\equiv p(x|y) \\
c_{21}(y; x) &\equiv p(y|x)
\end{aligned}
$$

From the product rule, for any choice of a, b,

$$
\begin{aligned}
f(a, b) &= m_1(a)\, c_{21}(b; a) \\
\rightarrow m_1(a) &= \frac{f(a, b)}{c_{21}(b; a)}, \text{ for } \textit{any } b \\
\text{similarly } m_2(b) &= \frac{f(a, b)}{c_{12}(a; b)}, \text{ for } \textit{any } a
\end{aligned}
$$

Now use the product rule for $p(x, y)$, replacing marginals:

$$
\begin{aligned}
f(x, y) &= m_1(x)\, c_{21}(y; x) \\
&= \frac{f(x, b)}{c_{21}(b; x)}\, c_{21}(y; x), \text{ for any } b \\
&= \frac{m_2(b) c_{12}(x; b)}{c_{21}(b; x)}\, c_{21}(y; x) \\
&= f(a, b)\frac{c_{12}(x; b)}{c_{12}(a; b)}\, \frac{c_{21}(y; x)}{c_{21}(b; x)}
\end{aligned}
$$

for any choice (a, b) (requires a *positivity condition*: support of joint = cartesian product of supports of marginals)

$$f(x, y) = f(a, b) \frac{c_{12}(x; b)}{c_{12}(a; b)} \frac{c_{21}(y; x)}{c_{21}(b; x)}$$

Here $f(a, b)$ is independent of (x, y), playing the role of a normalization constant for the remaining (x, y)-dependent factors

<div align="center">
Knowing all the conditionals
uniquely determines the joint
</div>

Uses of this result (and its generalizations):

- Pseudo-likelihood methods

- Complex graphical models—Markov random fields

- *Gibbs sampling*: Using conditionals to build a MH proposal distribution

Agenda

1 Joint from conditionals

2 Gibbs sampling

Gibbs sampling

Consider the MH algorithm for sampling from a 2-D distribution, $p(x, y)$, with proposal distribution $k(x', y'|x, y)$ for proposing a candidate new state (x', y') when the current state is (x, y)

The acceptance probability is $\alpha(x', y'|x, y) = \min r(x', y'|x, y), 1$ with

$$r(x', y'|x, y) = \frac{p(x', y')}{p(x, y)} \times \frac{k(x, y|x', y')}{k(x', y'|x, y)}$$

Suppose we update only x, by *sampling from the full conditional* $c_{12}(x; y) = p(x|y)$,

$$k(x', y'|x, y) = c_{12}(x'; y)\delta(y' - y)$$

The acceptance ratio is

$$r(x', y'|x, y) \;=\; \frac{p(x', y')}{p(x, y)} \times \frac{c_{12}(x; y')\delta(y - y')}{c_{12}(x'; y)\delta(y' - y)}$$

Accounting for $y' = y$ and using the product rule (being a bit cavalier with δs!),

$$
\begin{aligned}
r(x', y'|x, y) &= \frac{p(x', y')}{p(x, y)} \times \frac{c_{12}(x; y')\delta(y - y')}{c_{12}(x'; y)\delta(y' - y)} \\
&= \frac{p(x', y)}{p(x, y)} \times \frac{c_{12}(x; y)}{c_{12}(x'; y)} \\
&= \frac{p(y)c_{12}(x'; y)}{p(y)c_{12}(x; y)} \times \frac{c_{12}(x; y)}{c_{12}(x'; y)} \\
&= 1
\end{aligned}
$$

We always accept a proposal from a conditional!

If we only propose x updates, the chain is reducible $\rightarrow$ need to do one of these:

- Randomly pick which parameter to update at each step
- Cycle through all parameters in a fixed order

This also works for *blocks* of parameters in many-parameter problems
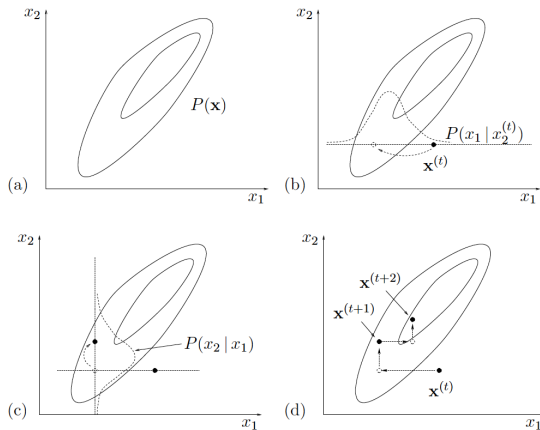
(a) The joint density $P(\mathbf{x})$ from which samples are required. (b) Starting from a state $\mathbf{x}^{(t)}$, x_1 is sampled from the conditional density $P(x_1 \mid x_2^{(t)})$. (c) A sample is then made from the conditional density $P(x_2 \mid x_1)$. (d) A couple of iterations of Gibbs sampling.

MacKay (2003)

Finding full conditionals

For $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$:

$$p(\theta_i | \boldsymbol{\theta}_{-i}) = \frac{p(\theta_1, \ldots, \theta_p)}{p(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p)}$$

Denominator doesn't depend on θ_i: The full conditional PDF for θ_i is just the *joint PDF*, considered only as a function of θ_i (and appropriately normalized)

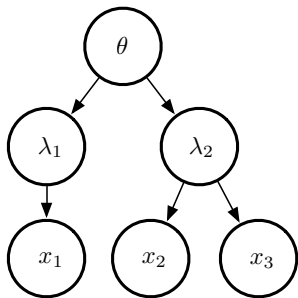For each parameter (or block of parameters)

- Write the joint PDF, ignoring any constants of proportionality

- Drop any factors that don't depend on θ_i

- Try to identify the remaining function as the kernel for a known PDF (some numerical methods relax this)

For graphical models, the DAG can guide identification of full conditionals—just use the factors from nodes that connect to the variable
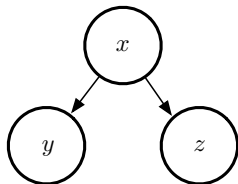
$$p(x, y, z) = p(x)p(y|x)p(z|x)$$
$$p(z|x, y) = \frac{p(x, y, z)}{p(x, y)} = p(z|x)$$

$$p(\theta, \lambda, x) = p(\theta)\, p(\lambda_1|\theta)\, p(\lambda_2|\theta)$$
$$\times p(x_1|\lambda_1)p(x_2|\lambda_2)p(x_3|\lambda_2)$$
$$p(\lambda_2|\ldots) \propto p(\lambda_2|\theta)\, p(x_2|\lambda_2)\, p(x_3|\lambda_2)$$
$$p(x_1|\ldots) = p(x_1|\lambda_1)$$