Jensen's Inequality for Conditional Expectations

by

S. M. Samuels

Purdue University

Department of Statistics

Division of Mathematical Science

Mimeo Series No. 313

January 1973

Jensen's Inequality for Conditional Expectations

by

S. M. Samuels

Purdue University

Let f be a real-valued Borel function, X and f(X) be integrable random variables defined on a probability space $(\Omega,\mathcal{F},P)$ and $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$. Jensen's inequality states that, if f is convex on an interval I containing the range of X, then

(1) $$Ef(X) \geq f(EX)$$

(where E denotes expectation); its generalization is

(2) $$E[f(X)|\mathcal{G}] \geq f[E(X|\mathcal{G})]$$

with probability one.

One application of the generalized Jensen's inequality is in martingale theory where it is used to show that "convex functions of martingales" and "convex non-decreasing functions of submartingales" are submartingales.

The usual proof of (1), e.g. Loève [6; p. 159], uses the fact that under the hypothesis there must exist a non-decreasing function, $m(\cdot)$, satisfying, for all x and y in I:

(3) $$f(x) - f(y) \geq m(y)\ (x - y)$$

(e.g. take m to be either the right or left hand derivative of f). Then, since EX must lie in I, we have

(4) $$f(X) - f(EX) \geq m(EX)\ (X - EX).$$

Take the expectations of both sides, giving

$$Ef(X) - f(EX) \geq m(EX)\ (EX - EX) = 0 \tag{5}$$

which proves (1).

Inequality (4) can be generalized to

$$f(X) - f[E(X|\mathcal{G})] \geq m[E(X|\mathcal{G})]\ [X - E(X|\mathcal{G})] \tag{6}$$

and taking the conditional expectations of both sides with respect to $\mathcal{G}$, would yield the analogue of (5), thereby proving (2) -- provided that the conditional expectations exist. If I can be bounded above or below then so can $m[E(X|\mathcal{G})]$ and there is no difficulty. But in general, the hypothesis is not sufficient to guarantee the existence of the mean of the right side of (6).

For example, take $f(x) = \exp|x|$, let Y be any symmetric random variable with $E\exp 2|Y| < \infty$, but $EY\exp|Y|$ failing to exist. Let Z be independent of Y with values 0 and 2 each with probability 1/2. Take X = YZ and $\mathcal{G}$ the σ-field generated by Y. Then the right side of (6) becomes

$$|Y|(Z-1)\exp|Y| \tag{7}$$

which has the same distribution as

$$Y\exp|Y|, \tag{8}$$

hence no mean. So while we are strongly tempted to say that the conditional expectation of (7) with respect to $\mathcal{G}$ is

$$(|Y|\exp|Y|)\ E(Z-1) = 0\ , \tag{9}$$

we may not do so.

Loève does not give a proof of (2), but merely asserts (p. 348) that it follows from (1) and the fact that $P\{X \geq Y\} = 1$ implies $P\{E(X|\mathcal{H}) \geq E(Y|\mathcal{H})\} = 1$ for any $\mathcal{H} \subset \mathcal{F}$. We leave this as an exercise for the reader: one which we have not been able to solve.

Feller [4; p. 214] mentions (2) without proof. Neveu [7; p. 122] mentions only the case $X \geq 0$, without proof as an easy generalization of (1), which it is since the interval I can be bounded below. Chung [2; p. 281] has a proof of (2) which is not based on (3) and which is not quite complete.

Doob [3; p. 33] shows that once the existence of regular conditional distributions is established, (2) can be obtained from (1) in an elementary way. Indeed Breiman [1; p. 80] assigns the proof of (2) as an exercise with the above as a hint.

I prefer to build a proof around (6) as follows:

Choose $a > 0$ and let

(10) $$A = A(a) = \{|E(X|\mathcal{G})| \leq a\} .$$

Then (6) is true with X replaced by XI_A -- unless 0 is not in I, in which case X should be replaced by $XI_A + bI_{A^c}$ for some b in I, and f(0) should be replaced by f(b) below. Now $m[E(XI_A|\mathcal{G})]$ is bounded, so we are justified in concluding that

(11) $$E[f(XI_A)|\mathcal{G}] \geq f[E(XI_A|\mathcal{G})] .$$

Because $A \in \mathcal{G}$, the left side of (11) is

(12) $$E[f(X)I_A + f(0)I_{A^c}|\mathcal{G}]$$
$$= E[f(X)|\mathcal{G}]I_A + f(0)I_{A^c} ,$$

while the right side is

(13) $$f[E(X|\mathcal{G})I_A] = f[E(X|\mathcal{G})]I_A + f(0)I_{A^c} .$$

Comparing (12) and (13) we see that, in effect, on A, the I_A's can be deleted from (11). Since $P(A) \to 1$ as $a \to \infty$, this completes the proof.

Recently, I noticed (6) in Hunt [5; p. 48] with the remark that (2) then follows immediately if X is bounded and "in general by a passage to the limit". So the preceding proof fills in the details omitted by Hunt. But note that the proof will not go through if (10) is replaced by $A = \{|X| \leq a\}$ since this set is not in $\mathcal{G}$.

REFERENCES

1. L. Breiman, Probability, Addison-Wesley, Reading, 1968.

2. K. L. Chung, A course in probability theory, Harcourt, Brace, and World, New York, 1968.

3. J. L. Doob, Stochastic processes, Wiley, New York, 1953.

4. W. Feller, An introduction to probability theory and its applications, vol. 2, 2nd. ed., Wiley, New York, 1971.

5. G. A. Hunt, Martingales et processus de Markov (in French), Dunod, Paris, 1966.

6. M. Loève, Probability theory, 3rd. ed., Van Nostrand, Princeton, 1963.

7. J. Neveu, Mathematical foundations of the calculus of probability, Holden-Day, San Francisco, 1965.